OPEN ACCESS

Journal of Low Power Electronics and Applications ISSN 2079-9268 www.mdpi.com/journal/jlpea

Review

Robust and Energy-Efficient Ultra-Low-Voltage Circuit Design under Timing Constraints in 65/45 nm CMOS

David Bol

ICTEAM institute, Université catholique de Louvain, Place du Levant 3, Louvain-la-Neuve, Belgium; E-Mail: david.bol@uclouvain.be; Tel.: +32-10472148; Fax: +32-10472598

Received: 23 November 2010; in revised form: 18 January 2011 / Accepted: 21 January 2011 / Published: 25 January 2011

Abstract: Ultra-low-voltage operation improves energy efficiency of logic circuits by a factor of $10\times$, at the expense of speed, which is acceptable for applications with low-to-medium performance requirements such as RFID, biomedical devices and wireless sensors. However, in 65/45 nm CMOS, variability and short-channel effects significantly harm robustness and timing closure of ultra-low-voltage circuits by reducing noise margins and jeopardizing gate delays. The consequent guardband on the supply voltage to meet a reasonable manufacturing yield potentially ruins energy efficiency. Moreover, high leakage currents in these technologies degrade energy efficiency in case of long stand-by periods. In this paper, we review recently published techniques to design robust and energy-efficient ultra-low-voltage circuits in 65/45 nm CMOS under relaxed yet strict timing constraints.

Keywords: digital CMOS circuits; ultra-low power; subthreshold logic; variability; leakage currents; yield

1. Introduction

Low power consumption is nowadays paramount for digital integrated circuits. High-performance chips such as multi-core processors for servers are power constrained by the die temperature limit and by both the cooling and electricity costs [1]. Portable applications such as smart phones obviously have an even tighter power budget for battery life concern, which drove innovation during the last decade in advanced power management techniques [2]. Besides these mainstream designs stands another chip category: ultra-low power circuits for applications such as RFID, biomedical devices and sensor

networks [3]. These application have in common a minute power budget as the circuits should operate either on tiny batteries (<1 cm³ [4]) or harvest energy from their environment [5]: from a few nW to hundreds of μ W. Fortunately, these applications feature low-to-medium speed requirements with target clock frequencies f_{target} from 10 kHz to 50 MHz, depending on the application and circuit topology. These relaxed speed constraints give room for power savings beyond simple frequency scaling or duty-cycled operation. Indeed, the supply voltage V_{dd} can be scaled down to reduce the energy required to switch on-chip capacitances at each clock cycle $E_{sw} \propto C_L V_{dd}^2$, as the associated delay penalty is acceptable given the relaxed cycle time T_{cycle} at low-to-medium f_{target} . Ultra-low-voltage operation is the extreme case where V_{dd} is aggressively scaled down to 0.3–0.5 V with potential energy savings above $10 \times$ when compared to nominal- V_{dd} operation at 1–1.2 V.

Figure 1. Maximum clock frequency f_{clk} and corresponding energy per cycle E_{cycle} at ultra-low voltage (SPICE simulations of an 8-bit multiplier [6] in 65 and 45 nm LP CMOS technologies, at 25 °C, nominal results).



Ultra-low-voltage (ULV) operation was proposed in the 1970s [7,8] and put back in light for digital circuits in 1999 at the *Int. Symp. on Low-Power Electronics and Design* [9]. When V_{dd} is reduced to or below the threshold voltage V_t , MOSFETs start to operate in near-threshold or subthreshold regime [8,9]. As the subthreshold I_{on} current is exponentially dependent on V_{dd} , the gate delay dramatically increases. As shown in Figure 1, it significantly reduces the maximum clock frequency for digital circuits. The resulting T_{cycle} penalty also has a detrimental side effect on the total energy per cycle composed by switching and leakages contributions: $E_{cycle} = E_{sw} + E_{leak}$. Indeed, the leakage energy increases when reaching subthreshold regime as it results from the integration of leakage power over T_{cycle} : $E_{leak} = V_{dd}I_{leak} \times T_{cycle}$. There is thus an optimum supply voltage V_{min} , which minimizes the energy to an E_{min} level [10], as depicted in Figure 1. The V_{min} level is often comprised between 0.25 and 0.5 V depending on the ratio between E_{sw} and E_{leak} , which varies accordingly to circuit parameters and technology characteristics through total leakage current I_{leak} , average switched capacitance per cycle C_L , gate delay and number of gates in the critical path [11]. This concept known as the minimum-energy point has received a lot of attention in the research community during the last decade [3,12] with numerous successful ULV chip implementations: microcontrollers for biomedical applications [13,14],

for wireless sensor nodes [5,15] as well as dedicated ASICs for biomedical applications [16,17], communication [18], image processing [19,20] or RFIDs [21].

Along with this ULV trend, new CMOS technology nodes have been introduced to maintain the historical increase in on-chip device density. Unfortunately in nanometer CMOS technologies, reaching E_{min} in practice raises important challenges because ULV operation magnifies the sensitivity of circuits against MOSFET variability, short-channel effects and leakage currents [6,12]. Several design solutions have recently been proposed to reliably operate nanometer CMOS logic circuits at ultra-low voltage under relaxed yet strict timing constraints: gate length upsize [6], process flavor [22] and MOSFET selection [23], circuit adaptation [22] and power gating [24]. In this paper, we provide for the first time a unified review of:

- The pitfalls of nanometer ULV circuits limiting their minimum V_{dd} for functional robustness and timing closure;
- The detrimental impact of stand-by periods on energy efficiency;
- The proposed techniques to overcome these limitations.

We specifically target 65 and 45 nm CMOS nodes as they share many characteristics: multiple process flavors, std- κ oxide/poly-Si gate stack and strained-Si, which give similar behaviors at ultra-low voltage as shown in Figure 1. To illustrate the findings, we combine chip measurements in 65 nm and simulation results in 45 nm. The results are based on the work carried out in this field at *UCLouvain* and more specifically on papers [6,22,24,25].

The paper is organized as follows. In Section 2, we recall the impact of CMOS technology scaling on ULV circuits and set up a framework for evaluating energy-efficiency under robustness and timing constraints. We then address the impact of these constraints on the minimum ultra-low V_{dd} : the speed limit and the functional limit in Sections 3 and 4, respectively. Existing solutions are also presented. Section 5 finally deals with the impact of stand-by periods on energy efficiency, given these constraints on minimum V_{dd} .

2. Energy Efficiency of ULV Circuits in Nanometer CMOS Technologies

CMOS technology scaling driven by Moore's law increases MOSFET density on a chip by a factor of two every 18–24 months. This is particularly useful for increasing the functionality of CMOS circuits without increasing die area and thereby by keeping manufacturing costs acceptable. It also boosts speed performances at each technology generation while reducing the energy required to perform a given function [26]. ULV circuits for ultra-low-power applications similarly benefit from these enhancements. Indeed, E_{sw} is effectively reduced thanks to lower on-chip capacitances C_L while gate delay at ultra-low voltage is improved thanks to a higher subthreshold I_{on} current resulting from the scaled V_t [12]. This leads to boosted speed performances at the minimum-energy point.

However, CMOS technology scaling also comes with severe drawbacks when reaching nanometer CMOS nodes: leakage currents including subthreshold I_{off} current and gate tunneling leakage [27], short-channel effects [27] and variability [28]. The impact of these nanometer MOSFET effects are magnified at ultra-low voltage [12]. A first consequence at device level is a reduction of the effective I_{on}/I_{off} ratio due to lower V_t values and short-channel increase of the subthreshold swing and of

the drain-induced barrier lowering (DIBL) effect. A major consequence at circuit level is on the minimum-energy level E_{min} which stopped scaling from 90 nm node and actually increases significantly at 45 nm node because of the combined effects of subthreshold swing, DIBL, gate leakage and statistical variability [29]. Fortunately, this E_{min} increase can be limited by choosing the optimum MOSFET (medium gate length and low V_t) within a versatile yet standard CMOS technology menu with good speed performances and negligible area penalty [23]. Moreover, fully-depleted Silicon-on-Insulator (SOI) technology can further save 60% of E_{min} [29] although this technology is not yet commercially available for industrial circuit design.

Beyond E_{min} scaling trend, a key challenge for ULV circuit design in nanometer CMOS technologies is to reliably operate at the corresponding supply voltage V_{min} of the minimum-energy point. Indeed, as shown in Figure 2(a), the minimum V_{dd} for a logic circuit is given by both timing and robustness constraints [6]. Speed must be sufficient to meet the timing constraint associated with the target frequency f_{target} of the application. The delay of the critical path has to be shorter than the cycle time $T_{cycle} = 1/f_{target}$. Moreover, even if safe timing closure is achieved, there is a functional limit V_{limit} on V_{dd} , which is independent from f_{target} . We set up a framework for evaluating energy efficiency under timing and robustness constraints in [6], illustrated in Figure 2. This framework shows that the f_{target} range of ultra-low-power applications can be divided into 3 regions for ULV circuits:

- R1 region where E_{sw} dominates and minimum V_{dd} is speed limited,
- R2 region where E_{leak} dominates and minimum V_{dd} is speed limited,
- R3 region where E_{leak} dominates and minimum V_{dd} is limited by functionality.

Figure 2. Minimum V_{dd} and energy per cycle E_{cycle} vs. the target frequency of the application f_{target} (SPICE simulations of an 8-bit multiplier [6] in 45 nm LP CMOS technology, at 25 °C, Monte-Carlo simulations addresses local variations through statistical extraction of worst-case speed and functional limits as well as mean I_{leak}).



Within this framework, it is obvious that E_{min} is only reached at one particular clock frequency f_{min} corresponding to a T_{cycle} equal to the critical path delay at V_{min} . f_{min} is in R1 region as E_{leak} accounts for 30% of E_{cycle} at the minimum-energy point. E_{min} can thus only be reached for one particular target

frequency. If f_{target} is higher than f_{min} , switching energy is wasted because V_{dd} is higher than V_{min} and, if f_{target} is below f_{min} , leakage energy is wasted because leakage power is integrated over a prohibitively long T_{cycle} . For example, E_{min} of an 8-bit multiplier in a 45 nm LP (Low-Power) CMOS technology is reached at $V_{min} = 0.38$ V and $f_{min} = 630$ kHz, as shown in Figure 2. E_{cycle} is within $E_{min} + 10\%$ between 200 kHz and 2 MHz. For f_{target} outside this range, Figure 2 shows that practical energy under robustness and timing constraints can thus significantly differ from E_{min} [6].

As the minimum-energy point $(V_{min}, f_{min}, E_{min})$ varies with technology generations according to technological characteristics [12], there is an optimum CMOS technology node for each f_{target} that minimizes E_{cycle} under timing and robustness constraints [12,30]. However, using an older CMOS technology is not optimum regarding die area and thus high-volume manufacturing costs. For this reason, we focus in this paper on techniques to reliably operate ULV logic at the minimum-energy point in 65/45 nm CMOS technologies.

Finally, let us introduce here that statistical MOSFET variations in nanometer CMOS technologies due to random dopant fluctuations, line edge roughness, oxide thickness variations, etc. have an important impact on energy efficiency. Indeed, these variability sources induce local within-die random V_t variations that exponentially affect subthreshold I_{on} and I_{off} [31]. The consequences at circuit level are [6]:

- A guardband on T_{cycle} or on the minimum V_{dd} for sufficient timing (parametric) yield because worst-case delay of critical paths has to be considered given its large statistical distribution;
- Increase in functional limit V_{limit} voltage to ensure sufficient functional yield for large chips;
- Increase in mean leakage I_{leak} because I_{leak} is a lognormal distribution (exponentially dependent on the normally-distributed V_t) with a mean value higher than the typical one.

It has further been reported that E_{sw} is also statistically distributed in nanometer ULV circuits because local gate delay distribution introduces random glitches with E_{sw} penalties [32]. However, for the sake of simplicity we do not consider this effect in this paper.

As shown in Figure 2, statistical variability leads to energy penalties. As local variations can hardly be compensated by circuit adaption due to their randomness from a MOSFET to another, it is important to consider statistical variability when designing ULV circuits in nanometer CMOS technologies. In next sections, we review the constraints on minimum V_{dd} to ensure circuit robustness given this high local variability in nanometer CMOS technologies.

3. Speed Limit on V_{dd}

3.1. Timing Constraint and the Minimum-Energy Point

The first constraint on minimum V_{dd} is a timing constraint on the critical path delay, which have to be lower than T_{cycle} given by the f_{target} of the application. Typical f_{target} for ultra-low-power circuits ranges from 10 kHz to 50 MHz. As explained in Section 2, minimum energy of ULV circuits can only be reached at a single clock frequency f_{min} . The challenge for the designers is thus to tune the circuit to make f_{min} meet the f_{target} of the application. This can be done by changing the V_t of MOSFETs in the circuit. Indeed, reducing V_t will exponentially boost speed performances at ultra-low voltage through an exponential increase of subthreshold I_{on} which can be expressed from the subthreshold drain current expression [6]:

$$I_{sub} = I_0 \times 10^{\frac{V_{gs} + \eta_{DIBL} V_{ds}}{S}} \times \left(1 - e^{\frac{-V_{ds}}{U_{th}}}\right)$$
(1)

where I_0 is a reference current proportional to the MOSFET size W/L_g that exponentially depends on V_t , S is the subthreshold swing, U_{th} the thermal voltage and η_{DIBL} the drain-induced barrier lowering (DIBL) factor. The impact of V_t reduction on E_{leak} at a given ultra-low V_{dd} is not significant [11]. Indeed, as I_{leak} is often dominated by subthreshold leakage in 65/45 nm CMOS, the exponential I_{leak} increase from a V_t reduction through I_0 parameter is compensated by the shorter critical path delay and thus T_{cycle} , as long as the MOSFETs remain in subthreshold regime:

$$E_{leak} = V_{dd} \times I_{leak} \times T_{cycle}$$

$$\propto V_{dd} \times I_0 \ 10^{\frac{\eta_{DIBL}V_{dd}}{S}} \times \frac{L_D C_L V_{dd}}{I_0 10^{\frac{(1+\eta_{DIBL})V_{dd}}{S}}}$$

$$\propto L_D C_L \ 10^{\frac{-V_{dd}}{S}} V_{dd}^2$$
(2)

where L_D is the logic depth (number of gates in the critical path) and gate delay is modeled with CV/I approximation. By changing I_0 reference current through V_t tuning, the voltage V_{min} and energy level E_{min} of the minimum-energy point are thus not modified while f_{min} can be exponentially tuned to make it corresponds to the f_{target} of the application.

Standard nanometer CMOS technologies feature a versatile technology menu with several process flavors targeting different applications: General-purpose (GP) also called generic (G) process targets high-performance applications with short gate delay and relaxed leakage constraints while low-power (LP) process targets portable applications with relaxed speed and tight leakage constraints [22]. GP flavor features short gate length, low V_t and thin oxide for maximizing I_{on} at nominal V_{dd} whereas LP flavor feature longer gate length and higher V_t for subthreshold leakage concern and thicker oxide for gate leakage concern. As a result, subthreshold current varies by several orders of magnitude between GP and LP flavor through I_0 reference current. Figure 3 illustrates this fact with the measured frequency of 65 nm ring oscillators in GP and LP flavors. At 0.35 V for example, GP flavor frequency (11 MHz) is $125 \times$ higher than LP flavor frequency (88 kHz). Notice that this speed difference is much higher than at nominal 1–1.2 V V_{dd} because of the exponential dependence of subthreshold current on V_t at ultra-low voltage.

We thus showed in [22] that process flavor selection can effectively be used to operate at the minimum-energy point (V_{min}, E_{min}) for a wide f_{target} range. LP flavor can be used for frequencies between 1 and 50 MHz. Moreover, nanometer CMOS technologies feature MOSFETs with two or three different V_t values within each flavor. Fine f_{min} tuning to meet f_{target} can thus further be achieved by proper V_t selection for the MOSFETs. As shown in Figure 3, moving from standard- V_t (SVT) to low- V_t (LVT) devices in 65 nm boosts frequency and thus f_{min} by factors of $5.6 \times$ and $1.75 \times$ in LP and GP flavors, respectively. The frequency difference between SVT and LVT is lower in GP flavor because at 0.35 V, GP MOSFETs operate more in the near-threshold regime ($V_t \approx 350 \text{ mV}$) than in subthreshold regime and the I_{on} dependence on V_t is not fully exponential. Let us mention here that the curve of energy vs. f_{target} is quite

flat in the vicinity of the minimum-energy point, as shown in Figure 2. Therefore, once a proper process flavor and V_t selection has been performed to bring f_{min} close to f_{target} , fine tuning of V_{dd} by a few tens of mV can be used for meeting exactly the timing constraint with negligible energy overhead [22].

Figure 3. Measured speed for different CMOS flavors and V_t 's (measurements of 251-stage ring oscillators with FO1 inverters [25] in 65 nm LP/GP CMOS technology, at 25 °C, mean frequency of 20 measured dies).



3.2. Timing Constraint and Process/Temperature Variations

As MOSFETs in ULV circuits operate in the near- or sub-threshold regime, not only their I_{off} but also their I_{on} current depend exponentially on V_t through I_0 parameter from Equation (1). Gate delay is thus very sensitive to V_t variations [31] coming either from local random variations, global process corners or temperature variations. The frequency distribution of a ring oscillator at 0.3 V on 20 dies in 65 nm LP CMOS is plotted in Figure 4 for three different operating temperatures. This figure also compares the results with simulations at extreme SS (Slow NMOS/Slow PMOS) and FF (Fast NMOS/Fast PMOS) process corners. At 25 °C, the frequency at SS corner is $6.5 \times$ lower than typical frequency, which induces a large T_{cycle} guardband to ensure sufficient timing (parametric) yield regarding the f_{target} timing constraint. However, we showed in [25] that the main concern regarding timing constraint in ULV circuits comes from low-temperature operation. Indeed, low-temperature operation dramatically reduces subthreshold I_{on} due to V_t increase and subthreshold swing reduction. The measured impact of a -40 °C operation on speed is a $8.5 \times$ delay increase at 0.3 V. The T_{cycle} guardband to ensure safe timing closure over the standard temperature range from -40 to +85 °C is thus more important than the guardband for handling global process variations. This obviously implies energy penalties as minimum V_{dd} for speed constraint has to be increased to handle low-temperature operation. The simulated combined effect of SS corner and -40 °C operation on speed is a degradation of gate delay by a factor of $40\times$. Notice that the speed of ULV circuits in GP flavor suffer less from process/temperature variations [25]. Indeed, their near-threshold operation limits the exponential dependence of gate delay on V_t .

Figure 4. Distribution of maximum frequency with process and temperature variations (measurements of 251-stage ring oscillators with FO1 inverters [25] in 65 nm LP CMOS technology with simulation results of global process corners, $L_g = 60$ nm).



Although local variations have a strong impact on gate delay as mentioned in Section 2, the consequence on speed performances is smaller than the effect of global process/temperature variations. Indeed, gate delay variability is averaged out over the high number of gates in critical paths [31] and the guardband on T_{cycle} is thus reduced. For example, simulations of the 8-bit multiplier from Section 2 in 45 nm LP at 0.3 V show a 3σ worst-case delay due to local variations $2.3 \times$ higher than the typical delay. This is further mitigated by the use of an upsized L_g required to improve noise margins, as will be explained in Section 4.1.

In order to limit T_{cycle} guardbands and E_{cycle} penalties due to process/temperature variations, adaptive techniques can be used. Assuming that clock frequency is fixed at f_{target} by the application, adaptation can be achieved through either V_{dd} scaling or body biasing. We showed in [22] that adaptive body biasing is potentially more energy-efficient than adaptive voltage scaling as it exactly compensates V_t variations while the circuit is constantly operated at V_{min} . However, adaptive body biasing raises practical implementation issues. Indeed, the body bias voltages to compensate process/temperature variations are quite high in 65/45 nm CMOS technologies due to the vanishing body effect in short-channel thin-oxide MOSFETs [22]. Measurement results of ring oscillators in 65 nm LP CMOS at 0.3 V show that forward body biasing by 300 mV only reduces gate delay by a factor of $5\times$, which is not sufficient when compared to the $8.5\times$ delay increase due to -40 °C operation only. Adaptive voltage scaling is more efficient to mitigate delay increase at low temperature. Figure 5 shows the measured minimum V_{dd} to keep the delay constant vs. temperature. A 75 mV V_{dd} boost is capable of fully compensating the -40 °C delay increase. This comes at the expense of a 50% E_{sw} penalty at such a low temperature.

Figure 5. Minimum V_{dd} for compensating temperature-induced speed variations (measurements of 251-stage ring oscillators with FO1 inverters [25] in 65 nm LP CMOS technology).



4. Functional Limits on V_{dd}

4.1. Noise Margin Constraint

When V_{dd} is reduced from 1–1.2V to ultra-low values, the I_{on} reduction leads to lower I_{on}/I_{off} ratio for subthreshold MOSFETs. The impact on ULV logic is not only a speed penalty but also a strong reduction of noise margins [33,34]. The vanishing noise margins can lead to soft errors due to a higher sensitivity to transient noise from crosstalk [35] or radiations [36]. In 65/45 nm CMOS, local V_t variations further degrade output logic levels of ULV circuits, which can even lead to hard "stuck-at" faults and thus a functional failure of several manufactured chips [33,37]. When the number of gates in a circuit increases, the probability of a hard fault increases and the minimum V_{dd} for functionality V_{limit} increases fast. Measurement of 90 nm ring oscillators in [38] show that the mean V_{limit} between 1 k and 1 M gates is increased from 0.2 to 0.35 V. Robust ULV operation can thus only be achieved by taking V_{limit} into account, which might significantly degrade energy efficiency if V_{limit} gets close to the minimum-energy voltage V_{min} .

A convenient way to evaluate noise margins of ULV logic was proposed in [33] with the simulation of a NAND gate cross-coupled with a NOR gate, similarly to SRAM static noise margin extraction. This benchmark circuit represents an infinite chain of alternating NAND/NOR gates, which is a worst case regarding noise margins as the NAND (resp. NOR) gate features the highest V_{ih} (resp. lowest V_{il}) level with the highest V_{ol} (resp. lowest V_{oh}) due to stacking of on transistors and parallel combination of off transistors [33]. Let us mention that precise noise margin extraction for a given circuit can be performed according to the method from [39] but for the sake of generality, we stick to the NAND/NOR method in this paper. Figure 6(a) shows the noise margins of ULV logic in 45 nm LP technology from statistical Monte-Carlo simulation with this benchmark circuit at 0.3 and 0.4 V. The wide noise margin distribution implies that many gates with low noise margins exhibit a high susceptibility to transient noise. The probability of gates with a negative noise margin is even not null, which means that hard errors might be encountered in a large ULV chip with many gates. At 0.4 V, noise margins are higher, which decreases the susceptibility to transient noise and the probability of hard errors but might also degrade energy efficiency.

Figure 6. Noise margin distribution of ULV logic (SPICE simulations of NAND2/NOR2 gates [33] in 45 nm LP CMOS technology, at 25 °C, 1 k Monte-Carlo runs).



In order to reliably operate at V_{min} , several techniques have been proposed to increase noise margins and thereby improve V_{limit} . In [33], the authors propose to upsize the transistor width of critical gates to improve their resilience to local V_t variations and thereby limit their worst-case noise margins. However, this comes at the cost of energy penalties due to high C_L and I_{leak} in the circuit [6]. A widespread technique for robust ULV operation consists in the restriction of logic gates from the standard-cell library [20]. Indeed, gates with large transistors stacks or a large number of parallel branches such as NAND/NOR gates with 4 inputs feature worse noise margins. Eliminating these cells for ULV operation is thus very efficient to improve circuit robustness at the cost of slight area overhead. Another solution was proposed in [40,41]: V_t balancing also called adaptive β ratio. This technique can be used to match the subthreshold current between NMOS and PMOS devices in case of "crossed" process corners with slow NMOS/fast PMOS or the opposite. Implemented with an adaptive body biasing scheme, this technique can only address global process variations as the area overhead for compensating statistical local variations would be unacceptable. Therefore, this technique significantly improves nominal noise margins but is not capable of mitigating local noise margin variations.

We showed in [6,12] that both the degradation of the subthreshold swing and the increase of DIBL factor due to short-channel effects in nanometer CMOS technologies threatens ULV circuit robustness by further degrading output logic levels and thereby increasing V_{limit} . Therefore, upsizing the gate length L_g of MOSFETs in ULV logic is able to significantly improve noise margins [6,12]. As shown in Figure 6(b), an upsize of the drawn L_g by 20 nm in 45 nm LP CMOS tightens noise margins distributions. The impact on functional die yield is computed for 0.3 V logic circuits with a varying number of gates N_{gates} from 1 k to 1000 M. We constrained the minimum noise margins to 20 mV for robustness against transient noise and extrapolated die yield with a simple model:

with η_{gate} the functional yield with a 20 mV noise margin constraint for the NAND2/NOR2 benchmark circuit (2 gates). Notice that this is quite a pessimistic assumption as it considers a logic circuit with only alternating NAND2/NOR2 gates. The resulting die yield is plotted for both 40 and 60 nm drawn L_g in Figure 7. It shows that the maximum number of logic gates in a circuit for 95% die yield is increased from 15 k at the minimum L_g to 4 M logic gates at the upsized L_g . This technique is thus very efficient to improve V_{limit} for robust ULV operation. Moreover, it does not bring energy penalty as the C_L increase due to an upsized L_g is significantly compensated by E_{leak} reduction thanks to reduced subthreshold swing, DIBL and variability [23,29].

Figure 7. Functional yield at 0.3 V with a 20 mV constraint on minimum noise margin (SPICE simulations of NAND2/NOR2 gates in 45 nm LP CMOS technology, at $25 \degree$ C, 50 k Monte-Carlo runs with 95% confidence interval plotted).



4.2. Hold Time Constraint

As ULV logic features a magnified sensitivity against local V_t variations, the statistical distribution of gate delay is quite large. It not only limits speed due to T_{cycle} guardband reported in Section 2 but also threatens functionality of ULV circuits due to potential hold time failures [37]. Indeed, local delay variations in the clock tree of ULV circuits might lead to large clock skew values between two branches of the clock tree and short logic paths might thus exhibit timing violations of hold constraint [42]. This is a critical point as hold time violations cannot be fixed by relaxing the clock frequency and generate a fault each time the path is triggered. Hold time failures thus sets another limit on the minimum V_{dd} for functionality V_{limit} .

We further showed in [25] that low-temperature operation magnifies the sensitivity of ULV gate delay to V_t variations because of the steeper subthreshold swing. Low temperature thus increases the probability of hold time violations due to this variability-induced clock skew. As this raises V_{limit} with potential energy penalties, low temperature has to be carefully addressed when checking the timing closure of hold constraints in ULV logic.

Although this problem can be addressed by upsizing the width or length of MOSFETs within the clock tree, it comes with E_{sw} penalty from C_L increase because the clock tree has a high activity factor. Another technique was recently proposed in [42]. The idea comes from the fact that RC interconnect delays are less important than gate delays at ultra-low voltage [43]. Therefore, the distributed buffering of a standard H-type clock tree at each level in the tree can be replaced by a single yet stronger bufferization stage at the clock root without incurring delay penalties within the clock tree. In this case, all leaf flip-flops in the tree share a common buffer stage, which can be composed of several series-connected buffers, and delay variations in this buffer thus do not introduce clock skew. This significantly reduces the probability of hold time failures. For circuits with more than a few kgates, a single bufferization stage might not be practical due to the prohibitively large dimension of buffers. In this case, the approach can be extended to a clock tree with a reduced depth of 2–4 buffer stages.

To validate this technique, we measured V_{limit} of two versions of a small logic circuit presented in [21]: one version with a standard distributed clock tree bufferization and a second with a single bufferization stage at the clock root: two large series-connected buffers. The V_{limit} histograms are plotted in Figure 8. The use of a single bufferization stage allows safe operation down to 0.23 V, whereas several dies of the circuit with standard distributed bufferization fail below 0.5 V. Let us recall here that ULV circuits in GP process flavor exhibit less delay variations as MOSFETs operate in near-threshold regime. They are thus less sensitive to variability-induced hold time violations.

Figure 8. V_{limit} distribution for two versions of a small logic circuit (measurements of an 8-bit AES coprocessor with 3500 gates [21] in 65 nm LP CMOS technology, at 25 °C. Hold time violations due to clock tree variability prevent from reliably operating below 0.5 V. The use of a clock tree with a single bufferization stage significantly improves V_{limit} thanks to mitigation of hold time violations.



5. Energy Efficiency and Stand-By Periods

Many ultra-low-power applications such as data logging in environmental [44] or biomedical [5] domains typically operate on a duty-cycled basis with long stand-by. Power consumed in stand-by mode

degrades energy efficiency by adding an overhead to the effective energy per active cycle E_{cycle} [45]. When assuming ideal clock gating for eliminating switching power during stand-by periods, the effective E_{cycle} can be expressed as [45]:

$$E_{cycle} = E_{act} + E_{stb}$$

= $E_{act} + P_{leak}T_{cycle} \times \frac{1 - \alpha_{duty}}{\alpha_{duty}}$ (4)

where α_{duty} is the duty cycle *i.e.*, the percentage of time that the circuit spends in active mode. As illustrated in Figure 9, an α_{duty} of 0.1% increases the effective E_{cycle} by a factor of 220× at the minimum-energy point.

Figure 9. Impact of stand-by periods on effective energy per cycle E_{cycle} (SPICE simulations of an 8-bit multiplier [6] 45 nm LP CMOS technology, at 25 °C.



To mitigate the E_{cycle} overhead, P_{leak} can be reduced either with an active-mode reduction technique or with a sleep-mode reduction technique [6]. Active-mode leakage reduction techniques relies on a V_t increase either globally in the whole circuit or selectively in gates from non-critical paths. At ultra-low voltage, a global V_t increase induces an exponential delay increase that requires a subsequent V_{dd} increase to maintain speed. If the V_t was already properly selected for making f_{min} meet the f_{target} of the application as proposed in Section 3, a global V_t assignment will make the minimum V_{dd} for speed deviate from V_{min} and in turn increase E_{cycle} [6]. Moreover, a selective V_t increase in non-critical paths is not efficient at ultra-low voltages because the exponential delay dependence on V_t due to MOSFET subthreshold operation limits the high- V_t assignment to a few logic gates in very short paths [22]. Besides V_t increase, serial operation was proposed in [46] to limit the number of gates and thereby reduce P_{leak} . This is an efficient technique which comes at the cost of more complex architectural design. In any case, active-mode leakage reduction techniques can only cut P_{leak} by a factor of $3-10 \times$ [6], which is not sufficient with α_{duty} values below 5%. **Figure 10.** Degradation of noise margins with sleep transistor sizing (SPICE simulations of an 8-bit multiplier for the leakage reduction [6] and NAND2/NOR2 circuit for noise margins ([33]) 45 nm LP CMOS technology, at 25 °C, sleep transistor width is normalized to the total width of parallel NMOS branches.



Therefore, a sleep-mode leakage reduction technique is preferred. Amongst them, power gating relies on the addition of a high- V_t sleep transistor to cut off the leakage path in sleep mode. The effective E_{cycle} can thus be expressed as [24]:

$$E_{cycle} = E_{act} + P_{sleep}T_{cycle} \times \frac{1 - \alpha_{duty}}{\alpha_{duty}} + \frac{E_{wake-up}}{N_{cycles}}$$
(5)

where P_{sleep} is the leakage power in stand-by mode, $E_{wake-up}$ the energy required to wake up from sleep mode and N_{cycles} the number of cycles in active mode between two stand-by periods in sleep mode. Notice that both wake-up and sleep-mode energies are amortized over N_{cucles} to calculate the effective energy per active cycle E_{cycle} . For sleep-mode energy, this is done through the $(1 - \alpha_{duty})/\alpha_{duty}$ term, which also corresponds to the ratio between cycles in sleep and active modes. Wake-up energy can usually be neglected when N_{cycles} is high (e.g., above 80 cycles in [24]). As in nominal- V_{dd} operation, the sleep transistor introduces a series resistance on the supply rails, which degrades ULV logic delay [45]. Sizing the sleep transistor thus results from a trade-off between large P_{sleep} reduction for narrow sleep transistors and small delay overhead for wide sleep transistors. Indeed, the delay overhead need a subsequent V_{dd} increase to meet the speed constraint with a subsequent E_{cycle} penalty [45]. Moreover, we showed in [24] that the series resistance of the sleep transistor also reduces noise margins of ULV logic. The consequence on V_{limit} for functional robustness is even worse than on minimum V_{dd} for speed. Figure 10 shows the impact of the sleep transistor sizing on P_{leak} reduction and the noise margin degradation at 0.35 V. A P_{leak} reduction by a factor of 100× reduces the noise margins by more than 50%, which makes ULV logic prone to functional failures. The impact of the sleep transistor on noise margin should thus carefully be addressed when designing a power-gated ULV circuit.

In order to limit this noise margin degradation, we showed in [24] that standard- V_t (SVT) MOSFETs with an upsized gate length should be preferred as they usually shows better subthreshold characteristics than high- V_t MOSFETs in 65/45 nm LP CMOS. As shown in Figure 10, this optimum sleep transistor in 45 nm LP CMOS degrades the noise margins by less than 20% for a P_{leak} reduction of 100×. These

results with optimum sleep transistor further show that power gating is much more efficient than dynamic reverse body biasing in ULV circuits with long stand-by periods as dynamic reverse body biasing only enable $10 \times P_{leak}$ reduction [6].

6. Conclusions

Ultra-low-voltage (ULV) operation between 0.3 and 0.5 V leads to minimum-energy consumption at the expense of speed for ultra-low-power applications. However, ensuring robust and energy-efficient ULV operation in nanometer CMOS technologies raises a number of design challenges due to high short-channel effects, leakage currents and variability of these technologies. In this paper, we reviewed these challenges and the potential circuit solutions, as summarized in Table 1.

Table 1. Design challenges for robust and energy-efficient ULV operation under timing constraints in 65/45 nm CMOS technologies.

Challenge	Circuit consequence	Preferred solution
Mismatch between f_{target} and f_{min}	E_{cycle} penalty	Process flavor & V_t selection
Operation at -40 °C	Delay increase— T_{cycle} guardband	Adaptive voltage scaling
Degraded noise margins	Soft and hard errors— V_{limit} increase	Upsized L_g & logic gate restriction
Variability-induced clock skew	Hold time violations— V_{limit} increase	Single-stage clock bufferization
Long stand-by periods	Effective E_{cycle} penalty	Power gating with opt. sleep transistor

First, we set up a general framework for analyzing energy efficiency under timing and robustness constraints for the whole range of target clock frequencies f_{target} in ultra-low-power applications. We specifically took the impact local V_t variations into account in this framework through statistical circuit simulation.

We then reported that the frequency of the minimum-energy point f_{min} can significantly differ from f_{target} with large E_{cycle} energy penalties. Process flavor and V_t selection in a versatile yet standard CMOS technology menu can be used to operate at the minimum-energy point under the timing constraint of the considered application, *i.e.*, make f_{min} meet f_{target} . We investigated the impact of global process/temperature variations on the timing constraint set by f_{target} . Low-temperature operation was shown to be a primary concern as it dramatically degrades delay and thereby involves large cycle time T_{cycle} guardbands. Adaptive voltage scaling was shown to be able to fix this at reasonable energy penalty.

We then analyzed how the minimum supply voltage for functionality V_{limit} is set by degraded noise margins and variability-induced clock skew. The first phenomenon induces soft errors due to increased noise sensitivity and even hard errors due to "stuck-at" faults. This can be fixed by gate length upsize and restriction of the logic gates within the standard-cell library to only low fan-in gates. The second phenomenon can lead to hold time violations and can be addressed by single-stage bufferization in the clock tree.

We finally analyzed the impact of stand-by periods on effective E_{cycle} . Application with low duty cycles need a leakage reduction technique to reduce leakage power in stand-by mode. Power-gating technique is preferred thanks to its high leakage power reduction capability. However, the addition of

the sleep transistor harms noise margins and thereby increases V_{limit} . This side effect can be effectively mitigated by the choice of an optimum sleep transistor.

Acknowledgements

Dr. Bol is with UCLouvain as a postdoctoral researcher from the National Foundation for Scientific Research (FNRS) of Belgium. Chip manufacturing was supported by the Walloon region of Belgium under TABLOID and E.USER projects. The author would like to thank C. Hocquet from UCLouvain for his precious help with chip measurements.

References

- 1. Bianchini, R.; Rajamony, R. Power and energy management for server systems. *Computer* **2004**, *37*, 68–76.
- Gammie, G.; Wang, A.; Mair, H.; Lagerquist, R.; Chau, M.; Royannez, P.; Gururajarao, S.; Ko, U. SmartReflex Power and Performance Management Technologies for 90 nm, 65 nm, and 45 nm Mobile Application Processors. *Proc. IEEE* 2010, *98*, 144–159.
- 3. Wang, A.; Calhoun, B.; Chandrakasan, A. *Sub-Threshold Design for Ultra-Low-Power Systems*; Springer: New York, NY, USA, 2006.
- Warneke, B.; Pister, K. An ultra-low energy microcontroller for Smart Dust wireless sensor networks. In *Proceedings of the 2004 IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA, 2004; pp. 316–317.
- Chen, G.; Fojtik, M.; Kim, D.; Fick, D.; Park, J.; Seok, M.; Chen, M.T.; Foo, Z.; Sylvester, D.; Blaauw, D. Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells. In *Proceedings of the 2010 IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA, 2010; pp. 288–289.
- Bol, D.; Ambroise, R.; Flandre, D.; Legat, J.D. Analysis and minimization of practical energy in 45nm subthreshold logic circuits. In *Proceedings of the 2008 IEEE International Conference on Computer Design*, Lake Tahoe, CA, USA, 2008; pp. 294–300.
- 7. Swanson, R.; Meindl, J. Ion-implanted complementary MOS transistors in low-voltage circuits. *IEEE J. Solid-State Circuits* **1972**, *7*, 146–153.
- 8. Vittoz, E.; Fellrath, J. CMOS analog integrated circuits based on weak inversion operations. *IEEE J. Solid-State Circuits* **1977**, *12*, 224–231.
- Soeleman, H.; Roy, K. Ultra-low power digital subthreshold logic circuits. In *Proceedings of the* 1999 ACM/IEEE International Symposium on Low Power Electronics and Design, San Diego, CA, USA, 1999; pp. 94–96.
- 10. Kao, J.; Miyazaki, M.; Chandrakasan, A. A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture. *IEEE J. Solid-State Circuits* **2002**, *37*, 1545–1554.
- 11. Calhoun, B.; Wang, A.; Chandrakasan, A. Modeling and sizing for minimum energy operation in subthreshold circuits. *IEEE J. Solid-State Circuit* **2005**, *40*, 1778–1786.
- 12. Bol, D.; Ambroise, R.; Flandre, D.; Legat, J.D. Interests and limitations of technology scaling for subthreshold logic. *IEEE Trans. VLSI Syst.* **2009**, *17*, 1508–1519.

- 13. Kwong, J.; Ramadass, Y.; Verma, N.; Chandrakasan, A. A 65 nm sub- V_t microcontroller with integrated SRAM and switched capacitor DC-DC converter. *IEEE J. Solid-State Circuits* 2009, 44, 115–126.
- Sridhara, S.; DiRenzo, M.; Lingam, S.; Lee, S.J.; Blazquez, R.; Maxey, J.; Ghanem, S.; Lee, Y.H.; Abdallah, R.; Singh, P.; Goe, M. Microwatt embedded processor platform for medical system-on-chip applications. In *Proceedings of the 2010 IEEE Symposium on VLSI Circuits* (*VLSIC*), Honolulu, HI, USA, 2010; pp. 15–16.
- Zhai, B.; Nazhandali, L.; Olson, J.; Reeves, A.; Minuth, M.; Helfand, R.; Pant, S.; Blaauw, D.; Austin, T. A 2.60 pJ/Inst Subthreshold Sensor Processor for Optimal Energy Efficiency. In *Proceedings of the 2006 IEEE Symposium on VLSI Circuits (VLSIC)*, Honolulu, HI, USA, 2006.
- 16. Kim, C.I.; Soeleman, H.; Roy, K. Ultra-low-power DLMS adaptive filter for hearing aid applications. *IEEE Trans. VLSI Syst.* **2003**, *11*, 1058–1067.
- Jocke, S.C.; Bolus, J.F.; Wooters, S.N.; Jurik, A.D.; Weaver, A.C.; Blalock, T.N.; Calhoun, B.H. A 2.6-μW sub-threshold mixed-signal ECG SoC. In *Proceedings of the 2009 IEEE Symposium on VLSI Circuits (VLSIC)*, Kyoto, Japan, 2009; pp. 60–61.
- Sze, V.; Chandrakasan, A. A 0.4-V UWB baseband processor. In *Proceedings of the 2007* ACM/IEEE International Symposium on Low Power Electronics and Design, Portland, OR, USA, 2007; pp. 262–267.
- 19. Sze, V.; Finchelstein, D.; Sinangil, M.; Chandrakasan, A. A 0.7-V 1.8-mW H.264/AVC 720p Video Decoder. *IEEE J. Solid-State Circuits* **2009**, *44*, 2943–2956.
- 20. Pu, Y.; Pineda de Gyvez, J.; Corporaal, H.; Ha, Y. An Ultra-Low-Energy Multi-Standard JPEG Co-Processor in 65 nm CMOS With Sub/Near Threshold Supply Voltage. *IEEE J. Solid-State Circuits* **2010**, *45*, 668–680.
- Hocquet, C.; Kamel, D.; Regazzoni, F.; Legat, J.D.; Flandre, D.; Bol, D.; Standaert, F.-X. Harvesting the potential of nano-CMOS for lightweight cryptography: An ultra-low-voltage 65 nm AES coprocessor for passive RFID tags. *J. Cryptogr. Eng.* 2011, *1*, 8 pp.
- Bol, D.; Flandre, D.; Legat, J.D. Technology flavor selection and adaptive techniques for timing-constrained 45 nm subthreshold circuits. In *Proceedings of the 2009 ACM/IEEE International Symposium on Low Power Electronics and Design*, San Francisco, CA, USA, 2009; pp. 21–26.
- 23. Bol, D.; Kamel, D.; Flandre, D.; Legat, J.D. Nanometer MOSFET effects on the minimum-energy point of 45 nm subthreshold logic. In *Proceedings of the 2009 ACM/IEEE International Symposium on Low Power Electronics and Design*, San Francisco, CA, USA, 2009; pp. 3–8.
- Bol, D.; Hocquet, C.; Flandre, D.; Legat, J.D. Robustness-aware sleep transistor engineering for power-gated nanometer subthreshold circuits. In *Proceedings of the 2010 IEEE International Symposium on Circuits and System*, Paris, France, 2010; pp. 1484–1487.
- 25. Bol, D.; Hocquet, C.; Flandre, D.; Legat, J.D. The detrimental impact of negative Celsius temperature on ultra-low-voltage CMOS logic. In *Proceedings of the 2010 European Solid-State Circuit Research Conference*, Seville, Spain, 2010; pp. 522–525.
- 26. Dennard, R.; Gaensslen, F.; Rideout, V.; Bassous, E.; LeBlanc, A. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J. Solid-State Circuits* **1974**, *9*, 256–268.

- 27. Roy, K.; Mukhopadhyay, S.; Mahmoodi-Meimand, H. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proc. IEEE* **2003**, *91*, 305–327.
- 28. Asenov, A.; Brown, A.; Davies, J.; Kaya, S.; Slavcheva, G. Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs. *IEEE Trans. Electron Device* **2003**, *50*, 1837–1852.
- 29. Bol, D.; Flandre, D.; Legat, J.D. Nanometer MOSFET Effects on the Minimum-Energy Point of Sub-45 nm Subthreshold Logic—Mitigation at Technology and Circuit Levelsc. *ACM Trans. Design Autom. Electr. Syst.* **2010**, *16*, 2–26.
- Seok, M.; Sylvester, D.; Blaauw, D. Optimal technology selection for minimizing energy and variability in low voltage applications. In *Proceedings of the 2008 ACM/IEEE International Symposium on Low Power Electronics and Design*, Bangalore, India, 2008; pp. 9–14.
- Zhai, B.; Hanson, S.; Blaauw, D.; Sylvester, D. Analysis and mitigation of variability in subthreshold design. In *Proceedings of the 2005 ACM/IEEE International Symposium on Low Power Electronics and Design*, San Diego, CA, USA, 2005; pp. 20–25.
- Kamel, D.; Hocquet, C.; Standaert, F.X.; Flandre, D.; Bol, D. Glitch-induced within-die variations of dynamic energy in voltage-scaled nano-CMOS circuits. In *Proceedings of the 2010 European Solid-State Circuit Research Conference*, Seville, Spain, 2010; pp. 518 –521.
- Kwong, J.; Chandrakasan, A. Variation-Driven Device Sizing for Minimum Energy Sub-threshold Circuits. In *Proceedings of the 2006 ACM/IEEE International Symposium on Low Power Electronics and Design*, Tegernsee, Germany, 2006; pp. 8–13.
- 34. Alioto, M. Understanding DC Behavior of Subthreshold CMOS Logic Through Closed-Form Analysis. *IEEE Trans. Circuits Syst. I* 2010, *57*, 1597–1607.
- 35. Nanua, M.; Blaauw, D. Investigating Crosstalk in Sub-Threshold Circuits. In *Proceedings of the* 8th International Symposium on Quality Electronic Design, San Jose, CA, USA, 2007; pp. 639–646.
- 36. Dhillon, Y.; Diril, A.; Chatterjee, A.; Singh, A. Analysis and optimization of nanometer CMOS circuits for soft-error tolerance. *IEEE Trans. VLSI Syst.* **2006**, *14*, 514–524.
- 37. Verma, N.; Kwong, J.; Chandrakasan, A. Nanometer MOSFET variation in minimum energy subthreshold circuits. *IEEE Trans. Electron Device* **2008**, *55*, 163–174.
- 38. Niiyama, T.; Piao, Z.; Ishida, K.; Murakata, M.; Takamiya, M.; Sakurai, T. Increasing minimum operating voltage (VDDmin) with number of CMOS logic gates and experimental verification with up to 1 Mega-stage ring oscillators. In *Proceedings of the 2008 ACM/IEEE International Symposium on Low Power Electronics and Design*, Bangalore, India, 2008; pp. 117–122.
- Pu, Y.; de Gyvez, J.; Corporaal, H.; Ha, Y. Statistical noise margin estimation for sub-threshold combinational circuits. In *Proceedings of the 13th Asia and South Pacific Design Automation Conference*, Seoul, Korea, 2008, pp. 176–179.
- 40. Pu, Y.; de Jesus Pineda de Gyvez, J.; Corporaal, H.; Ha, Y. Vt balancing and device sizing towards high yield of sub-threshold static logic gates. In *Proceedings of the 2007 ACM/IEEE International Symposium on Low Power Electronics and Design*, Portland, OR, USA, 2007; pp. 355–358.
- Hwang, M.E.; Roy, K. ABRM: Adaptive β-Ratio Modulation for Process-Tolerant Ultradynamic Voltage Scaling. *IEEE Trans. VLSI Syst.* 2010, *18*, 281–290.

- 42. Seok, M.; Blaauw, D.; Sylvester, D. Clock network design for ultra-low power applications. In *Proceedings of the 2010 ACM/IEEE International Symposium on Low Power Electronics and Design*, Austin, DX, USA, 2010; pp. 271–276.
- Hanson, S.; Zhai, B.; Seok, M.; Cline, B.; Zhou, K.; Singhal, M.; Minuth, M.; Olson, J.; Nazhandali, L.; Austin, T.; Sylvester, D.; Blaauw, D. Exploring Variability and Performance in a Sub-200-mV Processor. *IEEE J. Solid-State Circuits* 2008, 43, 881–891.
- Hanson, S.; Seok, M.; Lin, Y.S.; Foo, Z.Y.; Kim, D.; Lee, Y.; Liu, N.; Sylvester, D.; Blaauw, D. A Low-Voltage Processor for Sensing Applications With Picowatt Standby Mode. *IEEE J. Solid-State Circuits* 2009, 44, 1145–1155.
- 45. Seok, M.; Hanson, S.; Sylvester, D.; Blauw, D. Analysis and optimization of sleep modes in subthreshold circuit design. In *Proceedings of the 44th ACM/IEEE Design Automation Conference*, San Diego, CA, USA, 2007; pp. 604–699.
- 46. Khanna, S.; Calhoun, B.H. Serial sub-threshold circuits for ultra-low-power systems. In Proceedings of the 14th ACM/IEEE International Symposium on Low Power Electronics and Design, New York, NY, USA, 2009; pp. 27–32.

© 2011 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/.)