Analysis and Minimization of Practical Energy in 45nm Subthreshold Logic Circuits

David Bol, Renaud Ambroise, Denis Flandre and Jean-Didier Legat
Microelectronics laboratory, Université catholique de Louvain, Belgium
{*david.bol,renaud.ambroise,denis.flandre,jean-didier.legat*}*@uclouvain.be*

***Abstract*— Over the last decade, the design of ultra-low-power digital circuits in subthreshold regime has been driven by the quest for minimum energy per operation. In this contribution, we observe that operating at minimum-energy point is not straightforward as design constraints from real-life applications have an important impact on energy. Therefore, we introduce the alternative concept of practical energy, taking functional-yield and throughput constraints on minimum V_{dd} into account. In this context, we demonstrate for the first time the detrimental impact of DIBL on minimum V_{dd}.**

Practical energy gives a useful analysis framework of circuit optimization to reach minimum-energy point, while considering the throughput as an input variable dictated by the application. From simulation of a benchmark multiplier in 45nm technology, we find out that practical energy can be far higher than minimum energy point, in the case of low-throughput applications ($\approx$ 10-100 kOp/s) because of static leakage energy and robustness-limited minimum V_{dd}. With the proposed framework, we investigate the capability of conventional optimization techniques to make practical energy meet minimum energy point. Amongst these techniques, channel length upsize is shown to be more efficient than MTCMOS power gating, body biasing, V_t selection or device width upsize, as it increases robustness while simultaneously reducing static leakage energy. A small length upsize with low area overhead is shown to reduce practical energy at low throughput to less than $2.1\times$ the minimum energy level. At medium throughput, it even brings practical energy 30% lower than minimum energy level without optimization techniques.

I. Introduction

In the context of green computing, lowering the supply voltage V_{dd} is an efficient technique to drastically reduce dynamic energy per operation at the expense of increased delay [1]. However, as the delay increases, the execution time of the operation increases too. This implies a static-energy overhead, which results from the integration of the leakage power over the execution time of the operation. It has thus been shown that energy per operation can be minimized by operating at an optimal V_{dd}, which is often in subthreshold region. Over the last decade, minimum energy has become a popular research direction [2]-[8] for Ultra-Low-Power (ULP) applications such as RFID's, micro-sensor networks or implanted biomedical devices, which typically require low-to-medium data throughputs. Recently, minimum-energy subthreshold operation has also been proposed as an ULP mode in general-purpose portable systems with Dynamic Frequency Voltage Scaling (DFVS) [9]-[10].

The concept of minimum energy relies on 3 assumptions:

- optimal V_{dd} is high enough to ensure robust operation of the circuit;
- circuit delay is low enough to support required data throughput;
- circuit is powered down after completion of the operation with no energy consumption in sleep mode.

With these assumptions, the concept of minimum energy is quite theoretical. In real-life applications, V_{dd} cannot be scaled to arbitrarily-low values without raising robustness issues, especially with high process variability of nanoscale technologies [3]. Moreover, even with low data throughput requirement, ULP applications do have timing constraints. Yet very loose, they have to be taken into account. Finally, powering down the circuit is not straightforward in low-cost area/volume-constrained ULP applications. Circuit can be externally shut down but this requires extra off-chip components with cost and volume overheads and circuit state has to be saved in a memory with high energy overhead. On-chip circuit powering down can be achieved through multi-V_t (MTCMOS) technique with power-gating devices (sleep mode). Nevertheless, this technique adds design costs at architecture (management of active and sleep modes), gate (addition of special flip flops) and layout (separate routing of power grid for logic and flip flops) levels. Moreover, Seok *et al.* showed in [5] that sleep-mode energy cannot be overlooked and that the impact of power-gating devices on delay is higher in subthreshold regime.

In this contribution, we introduce the alternative concept of practical energy per operation, which takes robustness and throughput constraints into account. It brings a powerful analysis framework of energy efficiency, which treats data throughput as an input variable dictated by the application. We use 45nm predictive technology model from Arizona State University[1] [11] and a benchmark 8-bit RCA multiplier to show that practical energy in low-throughput applications can be far higher than minimum energy level because of high static energy and robustness-limited minimum operating V_{dd}. To solve these issues, a subthreshold processor in [12] uses MTCMOS power gating with device width upsize, whereas a subthreshold processor in [13] uses reverse body biasing and channel length upsize. It is thus not clear which technique is the most efficient. We use the proposed framework to carry

[1]Models available on-line at http://www.eas.asu.edu/~ptm.

978-1-4244-2658-4/08/$25.00 ©2008 IEEE

out for the first time an in-depth study of such techniques to pull practical energy toward minimum-energy point.

This paper is organized as follows. In Section II, the conventional concept of minimum energy is briefly reviewed and applied to the benchmark circuit. Constraints on minimum operating V_{dd} are presented in Section III. The impact on practical energy is then investigated in Section IV. Finally, minimization of practical energy is discussed in Section V.

II. Conventional concept of minimum energy

The energy per operation of a circuit is the sum of the dynamic energy E_{dyn} due to switched capacitances and the static energy E_{stat} due to the leakage current I_{leak} flowing through the devices. In the conventional approach, static energy is computed by the integration of I_{leak} during the actual execution of the operation i.e. over a time period equal to the delay of the circuit critical path. Energy is thus conventionally expressed as:

$$\begin{aligned} E_{conv} &= E_{dyn} + E_{stat} \\ E_{dyn} &= \frac{1}{2} N_{sw} C_L V_{dd}^2 \\ E_{stat} &= V_{dd} I_{leak} Del \end{aligned} \quad (1)$$

where N_{sw} is the number of switched nodes to perform the operation, C_L the typical node capacitance and Del the circuit delay. Minimum energy is often achieved when operating the circuit in MOSFET subthreshold region [2]-[8]. Subthreshold drain current is expressed as:

$$I_{sub} = I_0 \times 10^{\frac{V_{gs}+\eta V_{ds}}{S}} \times \left(1 - e^{\frac{-V_{ds}}{U_{th}}}\right) \quad (2)$$

where I_0 is a reference current proportional to W/L_{eff}, which depends exponentially on the threshold voltage V_t. S is the subthreshold swing, η the Drain-Induced Barrier Lowering (DIBL) coefficient and U_{th} the thermal voltage close to 26 mV at room temperature. At low V_{dd}, subthreshold leakage current is the dominating leakage source and static energy can be expressed as:

$$\begin{aligned} E_{stat} &= V_{dd} \times I_{leak} \times Del \\ &\propto V_{dd} \times I_0 \, 10^{\frac{\eta V_{dd}}{S}} \times \frac{L_D C_L V_{dd}}{I_0 10^{\frac{(1+\eta)V_{dd}}{S}}} \\ &\propto L_D C_L \, 10^{\frac{-V_{dd}}{S}} V_{dd}^2, \end{aligned} \quad (3)$$

where L_D is the logic depth.

Energy per operation of a benchmark 8-bit RCA multiplier has been extracted with this conventional approach from Spice simulation with 45nm PTM model ($T_{ox} = 1.1nm$, $L_{eff} = 17.5nm$ and nominal $V_{dd} = 1V$). A triple-V_t technology is considered with 0.27 low-, 0.37 std- and 0.46V high-V_t devices (extracted at $V_{ds} = 1V$). For the sake of generality, std-V_t devices at 25° are considered unless otherwise specified. Simulated energy per operation for a pseudo-random input pattern is shown vs. V_{dd} in Fig. 1. Under typical process corner, minimum energy happens at 0.3V optimal V_{dd}.

In nanoscale technologies, process variability cannot be overlooked especially when focusing on subthreshold design, as delay variability is magnified by the exponential dependence of I_{sub} on V_t [6]. Energy in the context of variability has to be computed with worst-case delay rather than typical delay. Through all this article, worst-case delay is statistically extracted from Monte-Carlo simulations with 99.9% confidence interval, i.e. 3σ tail of delay distribution. In Monte-Carlo simulations, we consider two sources of variability. First, random doping fluctuation is taken into account by modeling V_t of each device as an independant normally-distributed variable with a σ_{Vt} of 44mV for minimum-sized std-V_t devices, calculated with empirical expression from [14]. Second, variability of critical dimensions is considered through L_g variability. We model L_g variations as a normal distribution with $3\sigma/\mu$ equal to 20%. As L_g variations have a strong spatial correlation and the benchmark circuit is quite small, we consider only one normally-distributed L_g variable for all the devices of the circuit.

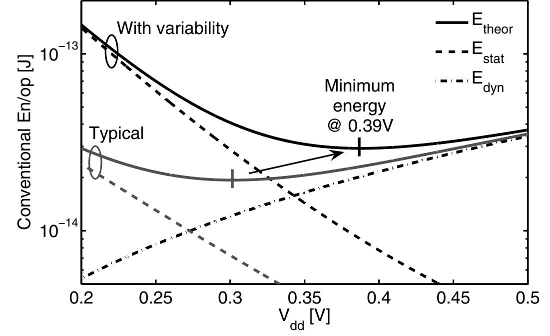

Fig. 1. Conventional energy per operation of a benchmark 8-bit RCA multiplier vs. V_{dd}, for 45nm technology with std-V_t devices. Variability is taken into account by considering statistically-extracted 3σ worst-case delay and mean I_{leak}.

When taking variability into account, static energy is far higher, resulting in a 90mV-higher optimal V_{dd} with a detrimental impact on minimum energy per operation, as studied in [7].

III. Constraints on minimum supply voltage

In previous section, minimum energy was computed assuming that V_{dd} can be reduced to an arbitrarily low value. In real-life applications, this is clearly not the case as a too low V_{dd} deteriorates functional yield and limits circuit throughput. In this section, we examine the minimum practical V_{dd} to meet both these constraints.

A. Functional yield constraint

Functional yield is an important issue in subthreshold circuits because subthreshold V_{dd} implies low I_{on}/I_{off} ratio and high current variability [8]. Current variability can in turn lead to bad output logic level of a gate, which would not be recognized as the correct logic level by the next gate. Some gates can thus exhibit functional failure, leading to bad functional yield of the subthreshold circuit. In [8], Kwong *et al.* proposes a very efficient method to extract functional

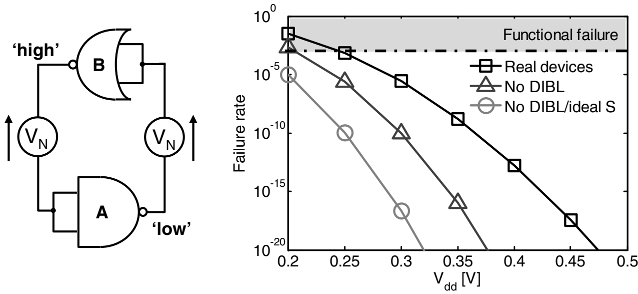

Fig. 2. Benchmark circuit to compute SNM distribution of logic gates (left) and resulting functional failure rate vs. V_{dd} (right). Functional yield is degraded by high subthreshold swing and DIBL.

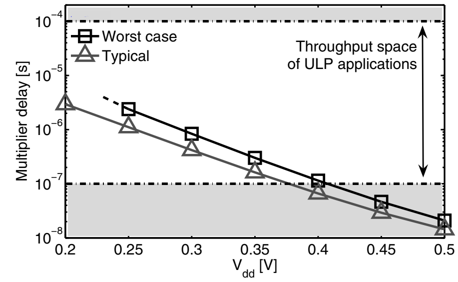

Fig. 3. Delay of the benchmark multiplier vs. V_{dd}. At 0.2V, worst-case delay is not plotted because 30% of the simulated occurences fail to operate correctly as suggested by high failure rate from Fig. 2.

yield of digital gates. This method, inherited from the extraction of SRAM-cell static noise margins (SNM), is based on a statistical computation of SNM of coupled gates depicted in Fig.2 (left) in the context of process variability. When SNM is negative it means that the output voltage V_{OL} of A is higher than the V_{IL} of B, and vice versa. The gates can thus not be operated at this supply voltage. The worst case consists of simulating a NAND gate, that has the highest V_{OL}, as A with a NOR gate, that has the lowest V_{IL}, as B. We use this method to simulate the SNM distribution with a 10k-point Monte-Carlo simulation. Functional yield is defined as the ratio of the occurrences having positive SNM. In this contribution, we arbitrarily specify a functional-yield constraint of 99.9% meaning that the worst-case SNM computed with 99.9% confidence interval (3σ tail) is positive.

Simulated functional failure rate is shown vs. V_{dd} in Fig. 2 (right). Failure rate dramatically increases when lowering V_{dd} below 0.3V. In order to ensure a functional yield of 99.9%, i.e. keeping functional failure rate lower than 0.1%, V_{dd} has to be kept at minimum 0.25V. This result could be surprising as circuits operating at less than 0.2V have been demonstrated in previous articles on subthreshold logic (e.g. [2]). However, these circuits were implemented on 0.15/ 0.18μm technologies. In 45nm technology, things are different. First, process variability is higher. Secondly, Hanson *et al.* show that SNM and thus functional yield are reduced with technology scaling due to increasing subthreshold swing S [15]. Finally, we observed for the first time that SNM and functional yield are considerably worsened by DIBL effect, which increases with technology scaling [16]. For the considered 45nm technology, we have the following values: S=96mV/dec and η=160mV/V. We carried out functional yield calculation with long-channel ideal S (77mV/dec) and without DIBL effect (η=0). Fig. 2 (right) shows that functional failure rate is considerably reduced under these conditions.

In order to validate these observations, let us get a qualitative insight on the impact of these effects on functional yield through SNM. Therefore, we examine the output level of a subthreshold inverter in DC operation, assuming an ideal high input level V_{dd}. The output is $V_{OL} \approx 0$. The pull-down NMOS in ON state has $V_{gs} = V_{dd}$ and $V_{ds} \approx 0$. The pull-up PMOS in OFF state has $\left|V_{gs}\right| = 0$ and $|V_{ds}| \approx V_{dd}$. DIBL effect acts as a systematic detrimental V_t mismatch between the ON device with low V_{ds} and the OFF device with $V_{ds} \approx V_{dd}$. The output voltage V_{OL} suffers from this and increases somewhat. We also consider variability through a detrimental V_t mismatch of 1σ. V_{OL} can be found by equating the currents of the pull-down and pull-up devices:

$$\begin{aligned} I_{sub,NMOS}|_{on} &= I_{sub,PMOS}|_{off} \\ I_0 \times 10^{\frac{V_{dd}-\sigma_{Vt}/2}{S}} \times \left(1 - e^{\frac{-V_{OL}}{U_{th}}}\right) &= I_0 \times 10^{\frac{\eta V_{dd}+\sigma_{Vt}/2}{S}} \\ 1 - e^{\frac{-V_{OL}}{U_{th}}} &= 10^{\frac{-((1-\eta)V_{dd}-\sigma_{Vt})}{S}} \\ V_{OL} &= -U_{th} \times ln\left(1 - 10^{-F_{SNM}}\right), \qquad (4) \\ with\ F_{SNM} &= \frac{(1-\eta)V_{dd} - \sigma_{Vt}}{S}. \end{aligned}$$

For very low V_{dd} values, V_{OL} depends on F_{SNM} as $10^{-F_{SNM}}$ is no longer negligible as compared to 1. Therefrom, an increase in S implies a deterioration of V_{OL} and thus SNM. Variability directly affects V_{OL} through σ_{Vt} in F_{SNM}. Moreover, it shows that the impact of DIBL effect is important through η that further lowers F_{SNM}. Under 0.3V, an η value of 200 mV/V has the same impact on V_{OL} that a V_t mismatch of 60mV or an increase of S by 25%. Subthreshold-swing improvement and DIBL mitigation are thus very important for robustness of subthreshold logic circuits in nanometer technologies.

B. Throughput constraint

Even ULP circuits have timing constraints, raised by the application. Circuits have to be fast enough to support the required data throughput, which ranges from medium to low values. Minimum V_{dd} thus depends on the application through the throughput requirement. In this contribution, we consider 10 kOp/s and 10 MOp/s, as the respective lower and upper bounds of the throughput region of interest for ULP applications. Fig. 3 shows the simulated delay of the benchmark multiplier, which depends exponentially on V_{dd} as subthreshold drain current does until 0.4V. Above 0.4V, increasing V_{dd} reduces less efficiently circuit delay because devices leave subthreshold regime.

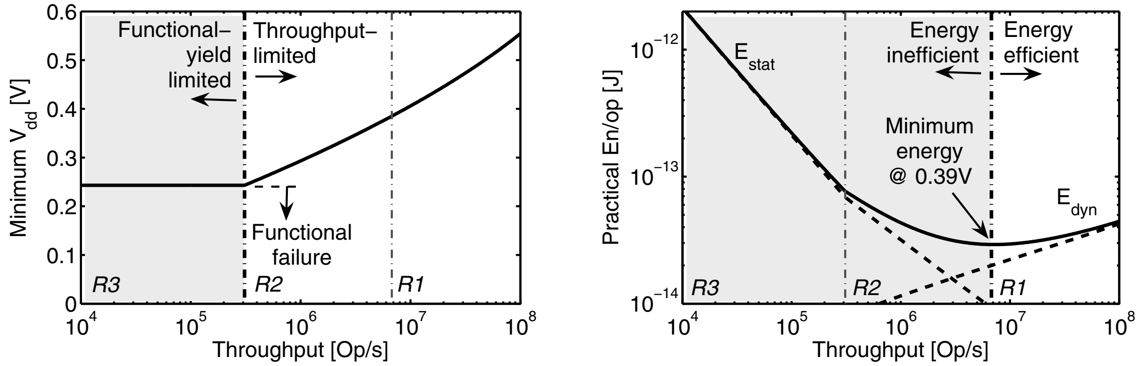

Fig. 4. Minimum operating V_{dd} (left) and corresponding practical energy per operation (right) of the benchmark multiplier. Throughput space can be divided into three regions: energy efficient ($R1$) and energy inefficient with either throughput ($R2$) or functional-yield ($R3$) V_{dd} limitation.

IV. Practical energy per operation

In conventional computation of minimum energy from Section II, the period to perform the operation was assumed to be equal to circuit delay. In practice, static energy due to I_{leak} is consumed over the whole period determined by the applied throughput. For the sake of generality, let us first assume in this section that the circuit does not have power-down feature to reduce static energy after completion of the operation. This will be addressed in Section V. Under this assumption, practical energy can be expressed as:

$$\begin{aligned} E_{pract} &= E_{dyn} + E_{stat} \\ E_{dyn} &= \frac{1}{2} N_{sw} C_L V_{dd}^2 \\ E_{stat} &= V_{dd} I_{leak} Per \\ &\propto V_{dd} \times I_0 10^{\frac{\eta V_{dd}}{S}} \times Per, \end{aligned} \quad (5)$$

where Per is the period to achieve the operation, i.e. the inverse of data throughput. From Eq. (5), it is clear that lowering V_{dd} is always beneficial to minimize practical energy. However, it was shown in Section III that V_{dd} cannot be reduced to arbitrarily-low values because it raises robustness and speed issues. Therefore, lowest practical energy is reached when operating at the minimum V_{dd} that meets both functional-yield and throughput constraints (just-in-time execution). Minimum V_{dd} is represented in Fig. 4 (left) vs. throughput. Notice once again that 3σ worst-case delay is chosen for verifying the throughput constraint. From this figure, the throughput space can be divided into two regions depending on the constraint that gives the highest limit for minimum V_{dd}. For the considered 45nm technology with std-V_t devices, minimum V_{dd} is limited by functional yield when throughput constraint is lower than 300 kOp/s.

Practical energy per operation of the benchmark multiplier under minimum V_{dd} from Fig. 4 (left) has been simulated. It is shown in Fig. 4 (right) vs. required throughput, dictated by the application. Throughput space can once more be divided into two new regions, depending on the dominating energy source. For the considered technology, above 7 MOp/s, dynamic energy dominates and the circuit is "energy efficient" because consumed energy actually contributes to perform the operation. However, when throughput is lower than 7 MOp/s, static energy dominates and the circuit is thus "energy inefficient". Moreover, for throughput below 300 kOp/s, static energy dramatically increases as V_{dd} cannot be reduced further because of functional-yield constraint. There is a minimum practical energy point, which corresponds exactly to the minimum energy point in the conventional approach of Section II. Indeed, when minimum V_{dd} is throughput-limited, E_{stat} from conventional and practical approaches are the same because $Del = Per$ from Eq. (1) and (5). Minimum practical energy happens when required throughput leads to V_{dd} value equal to optimal V_{dd} of minimum-energy point from Fig. 1. Minimum energy per operation thus happens in practice at only one particular data throughput. As throughput is determined by the application, it cannot be tuned by the designer, except with architectural modifications. This shows that minimum energy as described in previous articles on subthreshold logic is a theoretical concept. Minimum energy point can thus only be reached in practice by optimization techniques.

Let us summarize our observations. When looking at practical energy per operation, application throughput space can be divided into three regions:

- energy-efficient $R1$ region where dynamic energy dominates,
- energy-inefficient $R2$ region where static energy dominates and minimum V_{dd} is limited by throughput constraint,
- energy-inefficient $R3$ region where minimum V_{dd} is limited by functional-yield constraint.

As shown in Fig. 5, operating conditions do not change the picture but shift the throughput limits between these regions. A temperature increase implies a higher minimum V_{dd} to get sufficient functional yield, thereby extending $R3$. In combination with the increase of reference current I_0 with temperature (V_t lowering), this translates into a static energy overhead in $R3$. Static energy in $R2$ increases too due to subthreshold swing degradation.

When considering clock-gated circuits, the activity factor α_f is reduced. Fig. 5 shows that static energy is identical

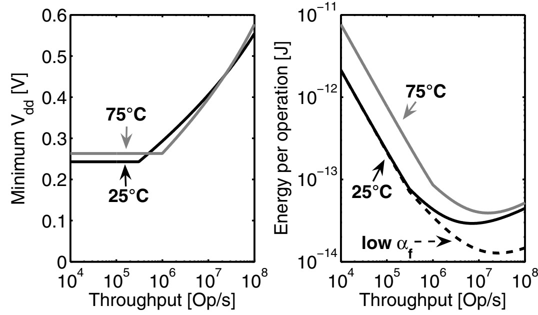

Fig. 5. Minimum operating V_{dd} (left) and corresponding practical energy per operation (right) for different operating temperatures and activity factors. α_f is divided by 3 for the dashed line to represent the impact of clock gating.

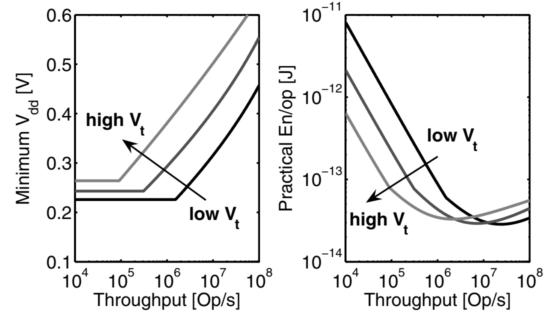

Fig. 6. Minimum operating V_{dd} (left) and corresponding practical energy per operation (right) for the different device types

whereas dynamic energy is lower. This extends $R2$ region and increases the difference between minimum energy level and practical energy in $R3$.

Pulling practical energy toward minimum energy point requires different achievements for each application throughput region. In $R1$ region, subthreshold current has to be increased in order to reduce delay and thus lower V_{dd} while meeting throughput constraint. In $R2$ region, subthreshold current has to be reduced at the expense of increased delay and thus V_{dd} to meet throughput constraint. In $R3$ region, subthreshold current has to be reduced and functional yield has to be increased in order to lower V_{dd}.

Looking at the throughput region of interest for ULP applications ($\approx$ 10 kOp/s - 10 MOp/s) in Fig. 4 (right), we see that circuit is nearly always energy inefficient because static energy dominates ($R2-R3$). Therefore, in next section, we address techniques to improve energy efficiency for $R2$ and $R3$ throughput regions.

V. Minimizing practical energy

In this section, we investigate the minimization of practical energy by circuit design techniques. Technology optimizations on one side and architectural optimizations on the other side are beyond the scope of this article. We first examine V_t selection in a multi-V_t technology. We then focus on run-time leakage reduction techniques: body biasing as in [13] and MTCMOS power gating as in [12]). We finally investigate device upsizing schemes to increase robustness: device width [12] and channel length [13] upsizes.

A. V_t selection

Nanoscale technologies often feature devices with various V_t values to trade off between speed and leakage power. Here, we consider a triple-V_t 45nm technology with 0.27 low-, 0.37 std- and 0.46V high-V_t devices. A higher V_t is achieved through increased channel doping, which results in lower subthreshold current. However, higher doping also leads to higher subthreshold swing because of higher depletion capacitance in the channel, and higher variability due to random doping fluctuation through σ_{V_t} calculated from [14].

Minimum V_{dd} is represented in Fig. 6 (left) for the 3 device types. Increasing V_t through channel doping, results in a higher minimum V_{dd} to meet throughput constraint as it increases circuit delay through lower subthreshold current. Moreover, high-V_t devices have a higher minimum V_{dd} to meet 99.9% functionnal-yield constraint because of higher variability and subthreshold swing. Notice that higher doping also decreases DIBL effect, but the impact of this reduction on SNM is negligible as compared to the impact of higher variability and subthreshold swing.

Simulated practical energy of the benchmark multiplier is shown in Fig. 6 (right). The use of low-V_t devices reduces dynamic energy as it allows to lower V_{dd} while meeting throughput constraint. Low-V_t devices are thus adapted to the upper part of throughput space of ULP applications. The use of high-V_t devices reduces static energy and is thus adapted to the lower part of the throughput space. Actually, a shift in V_t results in a shift of the practical energy curve vs. throughput. It thus extends the throughput range where practical energy is close to the minimum energy point. It is worth noticing that the lowest minimum energy is achieved with low-V_t devices as minimum energy is proportional to S^2 [15] and increases with variability [7].

B. Leakage reduction techniques

There are several ways to reduce subthreshold leakage current. Let us focus on 2 of them. The first technique we consider is the application of a Reverse Body Bias (RBB). This technique reduces MOSFET subthreshold leakage with cost overhead because it requires a triple-well process and the generation of 2 bias voltages ($-RBB$ and $V_{dd}+RBB$). In such low-voltage ULP applications, we consider a maximum affordable RBB of -0.6V. Minimum operating V_{dd} and corresponding practical energy per operation vs. the applied RBB are plotted in Fig. 7 in $R3$ (10 kOp/s) and $R2$ (1 MOp/s) throughput regions. In $R3$, RBB increases minimum V_{dd} because the higher subthreshold swing degrades functional yield. Nevertheless, practical energy is lowered thanks to leakage reduction but stays 25× higher than minimum energy level. In $R2$, RBB also increases minimumm V_{dd} as circuit delay increases due to reduced subthreshold current. There is only a small energy gain because what is win in terms of static energy is lost in dynamic energy because minimum V_{dd} increases. There is thus an optimum RBB in $R2$ close to -0.2V.

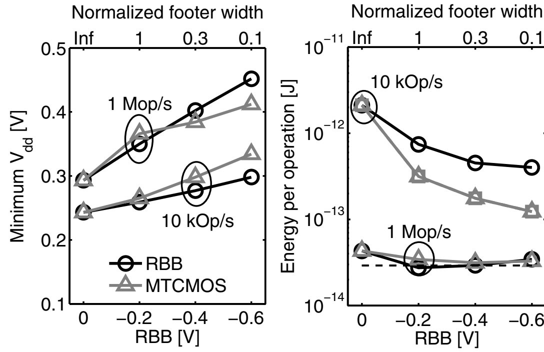

Fig. 7. Minimum operating V_{dd} (left) and corresponding practical energy per operation (right) with leakage reduction techniques in $R3$ (10 kOp/s) and $R2$ (1 MOp/s) regions (dashed line represents the level of minimum energy point)

The second technique we consider is MTCMOS power-gating technique: addition of a high-V_t NMOS footer to cut off leakage by entering sleep mode when the operation is completed. As explained in the introduction, this technique has a high design cost. As shown in Fig. 7 (left) for several footer widths (relative to the total NMOS width of the circuit), MTCMOS technique increases functional-yield limited minimum V_{dd} in $R3$ because it introduces a voltage shift of the circuit GND node, thereby reducing its effective V_{dd}. MTCMOS efficiently reduces practical energy, which tends to minimum energy level when decreasing footer width. However, there is lower bound on the footer width as it increases the GND voltage shift and could imply logic-level integrity issues at the interface between MTCMOS and non-MTCMOS (flip-flops) parts of the circuit. This results in a practical energy limit equal to 4× the level of minimum-energy point. In $R2$, minimum V_{dd} is raised because MTCMOS has a delay penalty. Notice that, in $R2$, minimum V_{dd} does not lead to optimal energy. Indeed raising somewhat V_{dd} reduces static energy in active-mode thanks to delay reduction. For the considered benchmark circuit, optimal V_{dd}, which is plotted in Fig. 7, is ±70mV higher than minimum V_{dd}. The corresponding optimal energy is reduced to a value close to minimum energy. However, notice that the energy of flip-flops, which cannot be power-gated, and the power-down/wake-up energies are not considered and would imply energy overheads.

C. Device upsizing schemes

Upsizing the devices reduces variability induced by random doping fluctuation proportional to $1/\sqrt{WL}$. The device dimensions that can be upsized at layout level are the width and the length. First, we consider an upsize of the device width to lower minimum operating V_{dd} through functional-yield increase, as proposed in [8]. Minimum V_{dd} and corresponding practical energy extracted from simulations are shown in Fig. 8, again for $R3$ and $R2$ throughput regions. In both regions, minimum V_{dd} is reduced by width upsize thanks to functional-yield improvement ($R3$) and delay variability mitigation ($R2$). However, the corresponding practical energy

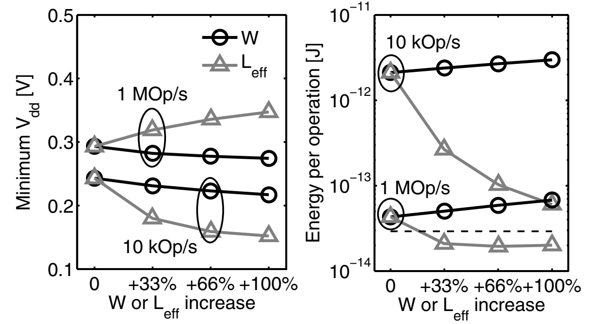

Fig. 8. Minimum operating V_{dd} (left) and corresponding practical energy per operation (right) with upsized devices in $R3$ (10 kOp/s) and $R2$ (1 MOp/s) regions (dashed line represents the level of minimum energy point)

TABLE I
COMPARISON OF OPTIMIZATION TECHNIQUES

	min V_{dd} in $R3$ [V]	E^*_{pract} in $R3$	min V_{dd} in $R2$ [V]	E^*_{pract} in $R2$
Standard	0.24	74.9	0.29	1.52
High V_t	0.26	22.5	0.38	1.21
RBB	0.30	25.5	0.35	1.18
MTCMOS	0.33	4.21	0.38	1.08
Length upsize	0.15	2.07	0.34	0.67

* Normalized to minimum energy without optimization techniques.

increases because the increase in switched capacitances and leakage currents outweighs the benefit of V_{dd} reduction.

Secondly, the channel length can be upsized to improve subthreshold operation as it not only mitigates variability but also improves subthreshold swing. Moreover, we observe that DIBL mitigation makes length upsize much more efficient in reducing minimum V_{dd} in $R3$ than width upsize, as shown in Fig. 8 (left). In addition, subthreshold swing improvement combined with this DIBL mitigation and V_t roll-off considerably lowers energy through leakage reduction. Doubling the effective channel length reduces practical energy by a factor 50, with small area overhead (only 17.5nm upsize). In $R2$, minimum V_{dd} increases as a length upsize has a detrimental impact on delay. At the same time, subthreshold swing improvement increases energy efficiency with small capacitance overhead. With these effects altogether, there is an optimal channel length, which allows to reduce practical energy by 55%, below the level of minimum energy point.

VI. CONCLUSION

In this contribution, we introduced a framework to analyze practical energy per operation of subthreshold circuits, taking functional yield and application throughput constraints into account. Throughput space can be divided in 3 regions depending on the dominating energy source and the limiting constraint on minimum V_{dd}. We showed that practical energy in 45nm technology can be far higher than minimum energy if required throughput is very low (10 kOp/s, $R3$ region).

We investigated practical energy minimization through reverse body biasing, MTCMOS power gating, V_t-selection and device upsize. We showed for the first time their impact on minimum operating V_{dd}. As shown in Table I, traditionally-used MTCMOS yields practical energy close to minimum energy with medium throughput (1 MOp/s, $R2$ region). However, it gives poor results at low throughputs with a practical energy more than 20× the minimum energy level. Channel length upsize is shown to be the most efficient technique thanks to subthreshold swing improvement, variability and DIBL mitigation. It reduces practical energy to less than 2.1× minimum energy at low throughputs and it even yields a practical energy 30% lower than minimum energy for medium throughputs, at small area cost. Additionally, we demonstrated for the first time the detrimental impact of DIBL on minimum operating V_{dd}.

ACKNOWLEDGMENTS

D. Bol and D. Flandre are research fellow and honorary senior research associate of the Fonds National de la Recherche Scientifique, respectively. R. Ambroise is with UCL thanks to a grant from the Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture.

REFERENCES

[1] H. Soeleman and K. Roy: "Ultra-low power digital subthreshold logic circuits", in *Proc. Int. Symp. Low-Power Electronics Des.*, 1999, pp. 94-96.

[2] J.T. Kao, M. Masayuki and A.P. Chandrakasan: "A 175-mV multiply-accumulate unit using an adaptative supply voltage and body bias architecture", in *IEEE J. Solid-State Circuits*, vol. 37 (11), pp. 1545-1554, 2002.

[3] N. Verma, J. Kwong and A. P. Chandrakasan: "Nanometer MOSFET variation in minimum energy subthreshold circuits", in *IEEE TED*, vol. 55(1), pp. 163-174, 2008.

[4] B. H. Calhoun, A. Wang and A. P. Chandrakasan: "Modeling and sizing for minimum energy operation in subthreshold circuits", in *IEEE J. Solid-State Circuits*, vol. 40 (9), pp. 1778-1786, 2005.

[5] M. Seok, S. Hanson, D. Sylvester and D. Blauw: "Analysis and optimization of sleep modes in subthreshold circuit design," in *Proc. Des. Autom. Conf.*, 2007, pp. 694-699.

[6] B. Zhai, S. Hanson, D. Blauw and D. Sylvester: "Analysis and mitigation of variability in subthreshold design," in *Proc. Int. Symp. Low-Power Electronics Des.*, 2005, pp. 20-25.

[7] S. Hanson, B. Zhai, D. Blaauw, D. Sylvester, A. Bryant and X. Wang: "Energy optimality and variability in subthreshold design", in *Proc. Int. Symp. Low-Power Electronics Des.*, 2006, pp. 363-365.

[8] J. Kwong and A.P. Chandrakasan: "Variation-driven device sizing for minimum energy sub-threshold circuits," in *Proc. Int. Symp. Low-Power Electronics Des.*, 2006, pp. 8-13.

[9] B. Zhai, D. Blaauw, D. Sylvester and K. Flautner: "The limit of dynamic voltage scaling and insomniac dynamic voltage scaling," in *IEEE Trans. VLSI Systems*, 13 (11), pp. 1239-1252, 2005.

[10] B.H. Calhoun and A.P. Chandrakasan: "Ultra-dynamic voltage scaling (UVDS) using sub-threshold operation and local voltage dithering", in *IEEE J. Solid-State Circuits*, vol. 41 (1), pp. 238-245, 2005.

[11] W. Zhao and Y. Cao: "New generation of predictive technology model for sub-45nm early design exploration," in *IEEE Trans. Electron Devices*, vol. 53 (11), pp. 2816-2823, 2006.

[12] J. Kwong, Y. Ramadass, N. Verma, M. Koesler, K. Huber, H. Moormann and A. Chandrakasan: "A 65nm sub-V_t microcontroller with integrated SRAM and switched-capacitor DC-DC converter," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2008, pp. 20-21.

[13] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhla, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester and D. Blaauw: "Exploring variability and performance in a sub-200-mV processor," in *IEEE J. Solid-State Circuits*, vol. 43 (4), pp. 881-891, 2008.

[14] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya and G. Slavcheva: "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," in *IEEE Trans. Electron Devices*, vol. 50 (9), pp. 1837-1852, 2003.

[15] S. Hanson, M. Seok, D. Sylvester and D. Blauw: "Nanometer device scaling in subthreshold logic and SRAM," in *IEEE Trans. Electron Devices*, vol. 55 (1), pp. 175-185, 2008.

[16] D. Bol, R. Ambroise, D. Flandre and J.-D. Legat: "Impact of technology scaling on digital subthreshold circuits," in *Proc. IEEE Comp. Soc. Annual Symp. VLSI*, pp. 179-184, 2008.