Roadmap for nanometer ultra-low-power digital circuits based on sub/near-threshold CMOS logic

D. Bol, D. Flandre and J.-D. Legat
Microelectronics laboratory, Université catholique de Louvain, Louvain-la-Neuve, Belgium
{david.bol,denis.flandre,jean-didier.legat}@uclouvain.be

Extracted from [6]: D. Bol, "Pushing Ultra-Low-Power Digital Circuits into the Nanometer Era", Ph.D dissertation, Université catholique de Louvain, Dec. 2008.

For a decade, low-power design has been part of mainstream semiconductor R&D activities, while ultra-low-power (ULP) design remained dedicated to niche markets of particular applications such as sensor networks, RFIDs and biomedical devices, where speed performances are not critical. However, today's 65/45nm CMOS technologies offer so compact circuits and high device performance that they give the opportunity for the designers to trade Silicon die area for energy efficiency, while maintaining sufficient speed. As a consequence, ultra-low-power design is now an emerging solution for energy-efficient mainstream electronics applications both in the consumer and professional markets. An important trend for designing ultra-low-power circutis is to aggressively scale the supply voltage V_{dd} to sub or near-threshold regime. In this document, we summarize the conclusions from the work carried out at the Université catholique de Louvain in this field, to set up a possible roadmap for sub/near-threshold CMOS logic in nanometer technologies.

1. Introduction

Scaling down the supply voltage V_{dd} of CMOS digital circuits is a straightforward technique to reduce the energy per operation E_{op} as its switching component E_{sw} is quadratically reduced. However, as shown in Fig. 1 for a benchmark multiplier in 0.13µm technology, it also results in a delay overhead and thus a reduction of the maximum operating frequency. When considering very low target frequencies f_{target} for low-performance applications (RFIDs, biomedical, sensor networks), V_{dd} can aggressively be scaled down. The delay increase results in long execution time of the operation. The consequence is a dramatic increase in leakage energy E_{leak}, which comes from the integration of leakage power over the execution time. This leads to a particular V_{dd}/f_{target} point of operation that minimizes energy consumption [1]. This minimum-energy point is denoted by V_{min}, f_{min}, E_{min}. It often occurs for supply voltages close or below the threshold voltage ($V_{min} \leq V_t$), leading to sub/near-threshold operation of MOSFETs. This decreases the I_{on}/I_{off} ratio and magnifies the sensitivity against process and temperature variations [2].

Looking at the technology nodes considered in recently-published subthreshold circuits, we can see that designers tend to naturally follow the technology scaling trend dictated by Moore's law [2]. A possible reason for this evolution is the reduction of manufacturing costs for high-volume products, where the cost of raw material i.e. Silicon wafer dominates. As shown in Fig. 1, the migration to nanometer technology nodes significantly reduces E_{sw} in sub/near-threshold circuits thanks to lower capacitances and minimum V_{dd} for achieving a given f_{target}. Unfortunately, it also dramatically increases E_{leak} because of higher leakage currents and minimum Vdd for ensuring functionality. This leads to two major issues when designing subthreshold circuits in nanometer technologies. First, the minimum energy level E_{min} increases by more than 50% when migrating from 130/90nm to 65/45nm technologies [3]. Second, the minimum-energy point is shifted towards higher f_{target} [2]. As shown in Fig. 1, f_{min} is shifted from the f_{target} range of low-performance (RFIDs, biomedical, sensor networks) to mid-performance (consumer/professional portable devices) ultra-low-power (ULP) applications between 130 and 45nm nodes.

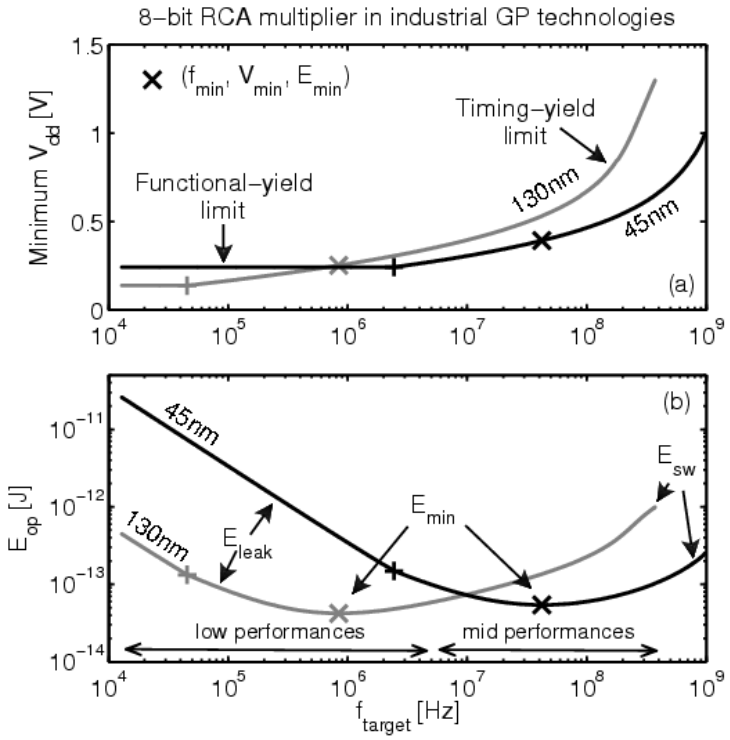

Fig. 1 – Simulated (a) minimum V_{dd} and (b) corresponding energy per operation E_{op} vs. the target frequency f_{target} in GP technologies (Spice simulations with foundry models) [1]

In this document, we summarize the conclusions from the work carried out at the Université catholique de Louvain in this field [1-6], to derive technology and circuit specifications for optimum CMOS sub/near-threshold circuits operating at the minimum-energy point. The explicit target is to fix the aforementioned issues by:

- keeping E_{min} level under control;
- reaching E_{min} under robustness and timing constraints for a wide performance range.

We then propose a possible roadmap for meeting these specifications in successive technology generations.

2. TECHNOLOGY/CIRCUIT SPECIFICATIONS FOR OPTIMUM SUB/NEAR-THRESHOLD OPERATION

As illustrated in Fig. 2, we propose to deal with the aforementioned major issues in two separate steps: reducing E_{min} as much as possible with technology/device optimizations and then reaching E_{min} in practice by relying on circuit techniques. The resulting technological targets as well as the circuit techniques requirement are detailed hereafter.

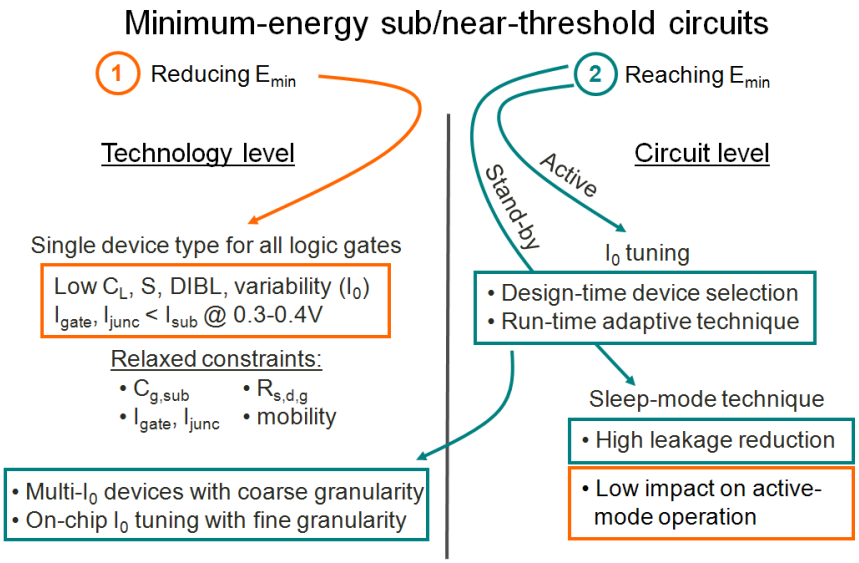

Fig. 2 – Technology/circuit specifications for optimum minimum-energy sub/near-threshold operation [6]

Reducing E_{min}

In [3], we showed that there are multiple effects in nanometer technologies that contribute to E_{min}. Limiting these effects leads to five key targets for optimum subthreshold MOSFET design at technology/device level in order to reduce E_{min}:

- low subthreshold swing S,
- low DIBL effect,
- low variability of subthreshold reference current I_0[†] (mainly from V_t variability),
- low mean load capacitance C_L including the device parasitic capacitances,
- gate and junction leakages, I_{gate} and I_{junc}, below the level of subthreshold leakage I_{sub}.

These technological targets are quite general. Indeed, they are also valid when designing devices for nominal-V_{dd} super-threshold circuits, beyond the scope of ULP applications. However, in ULP sub/near-threshold circuits, their importance is magnified. The good point is that several technological constraints of nominal-V_{dd} circuits are relaxed when considering subthreshold circuits. First, the intrinsic gate capacitance in subthreshold regime $C_{g,sub}$ is less important due to the addition of the channel depletion capacitance C_{dep} in series with the oxide capacitance C_{ox}. Therefore, $C_{g,sub}$ contributes less to C_L than in nominal-V_{dd} circuits [2-3]. It can thus be increased to achieve the targets of S, DIBL and variability minimization. Second, as the on-state subthreshold current is quite low, the equivalent channel resistance is large even in on-state. Therefore, the parasitic resistances associated to the device accesses R_s, R_d and R_g are proportionally less important. They can thus be increased without speed penalty in order to meet the other technological targets. Third, the subthreshold reference current I_0 exponentially depends on V_t and linearly on the carrier mobility. Mobility degradation can thus be tolerated as it is easily compensated by slight V_t reduction. Finally, I_{gate} and I_{junc} leakage currents do not have to be minimized. Technology designers should only prevent them to become higher than the subthreshold leakage. These relaxed constraints give space for device optimization to meet the five key targets. Notice that a single device type common to all logic gates can be used as dual-V_t/T_{ox}/L_g assignments are not practical in subthreshold logic circuits [1]. Only SRAM circuits may require different devices for leakage concern. This may possibly reduce the number of masks and process steps, and thereby save associated manufacturing costs.

[†] I_0 subthreshold reference current comprises all the bias-independent terms from subthreshold current expression [2]: $I_{sub}=I_0\times10^{(Vgs-\eta Vds/S)}\times(1-e^{-Vds/Uth})$, where η is the DIBL coefficient, S the subthreshold swing and U_{th} the thermal voltage. I_0 thus depends linearly on the device dimensions W/L and exponentially on $-V_t/S$.

Reaching E_{min}

Making the minimum-energy point meet the application target frequency, during active periods, requires I_0 tuning capability [1,6]. First, at design time, circuit designers should choose the technology with an I_0 value that brings minimum-energy point in the performance range of the target application i.e. $f_{min} \approx f_{target}$. This means that the technology should be versatile and come with multi-I_0 devices with a coarse granularity (e.g. three or four I_0 values in a wide range from 10 pA/µm to 10 nA/µm). Moreover, a low-cost circuit technique is required for post-Silicon I_0 tuning: at test time for compensating modeling errors or extrinsic global process variations, and/or at run time for compensating device aging, variations of the environment temperature or a dynamically-varying workload [1,6]. This implies that such a technique should be enabled at technology level for fine-grain tuning (a smaller I_0 range, e.g between 0.2× and 5× the nominal value). When the application features stand-by periods, a sleep-mode technique should be used at circuit level with strong leakage-reduction capability when in sleep mode. This technique should feature a low impact on delay and robustness when in active mode to avoid ruining the E_{min} level [6].

3. A POSSIBLE TECHNOLOGY/CIRCUIT ROADMAP FOR NANOMETER ULTRA-LOW-POWER DIGITAL CIRCUITS

We now would like to give the authors' personal view on the valuable technologies and circuit techniques to achieve the specifications we derived in previous section. As the application spectrum of ULP circuits is quite wide, we divided it into two categories depending on their performance requirements for the applications: low performances (f_{target} = 10 kHz to 5 MHz) for applications such as RFID tags, biomedical devices and sensor networks, and mid performances (f_{target} = 5 to 100 MHz) for portable consumer/professional applications. As illustrated in Fig. 3, we also divided the roadmap in three groups related to technology nodes: deep-submicron 130/90nm, present nanometer 65/45nm and future nanometer 32/22nm nodes.

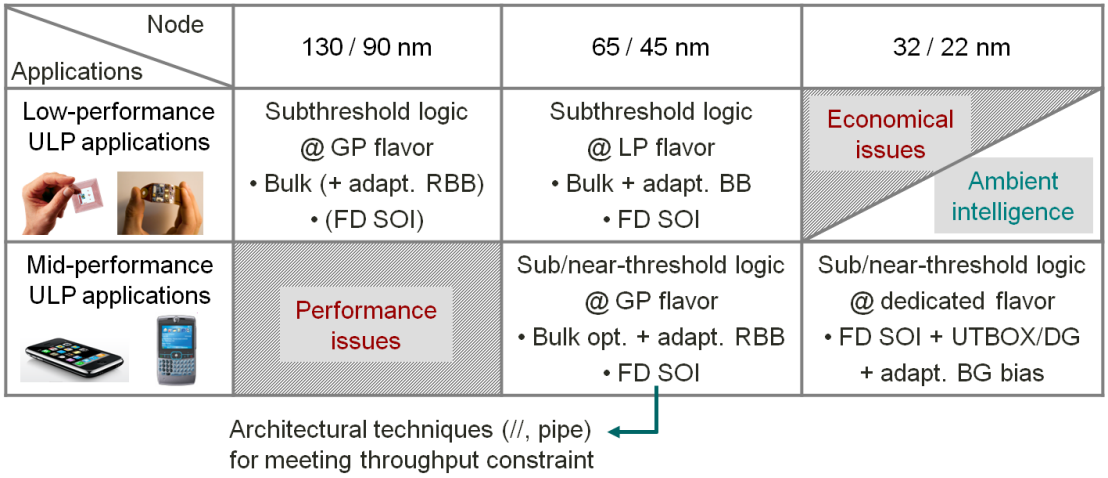

Fig. 3 – A possible technology/circuit roadmap for nanometer ultra-low-power digital circuits [6]

Low-performance ULP applications

In [2], we showed that subthreshold CMOS logic at deep-submicron 130/90nm nodes is well adapted to low-performance ULP applications even in bulk technology. At these nodes, foundries usually provide only a GP technology flavor. Fortunately, this is the most appropriate for providing I_0 values that make minimum-energy point meet target throughputs in the range of ULP applications. Fully-depleted SOI technology or adaptive body biasing (ABB) technique are interesting options for further energy savings by respectively improving MOSFET subthreshold characteristics (technology targets for reducing E_{min}) and removing design margins on V_{dd} through I_0 adaptation (circuit techniques requirements for reaching E_{min}). However, efficient subthreshold circuits have already been demonstrated at these nodes without these options [8], which shows that they are not compulsory.

When it comes to nanometer technologies at present 65/45nm nodes, only a low-power (LP) technology flavor features I_0 values compatible with the f_{target} range of low-performance ULP applications [1]. Without such an LP flavor, the benefit of die area reduction is waisted by an energy increase from 130/90 to 65/45nm nodes and there is thus no interest in migrating to nanometer technologies for energy concern [2]. Moreover, in order to avoid prohibitive design margins, which may prevent from reaching E_{min} in practice, an ABB technique is needed for test- or run-time circuit adaptation. Notice that reverse ABB is more energy-efficient than forward ABB [1] but more difficult to setup in practice, which implies higher design and Silicon costs. For those applications, the higher cost of reverse ABB with its possible detrimental impact on form factor might not be supported and forward ABB could thus be preferable. This is confirmed by a recent subthreshold design in [9] using forward ABB. Alternatively, an undoped-channel ultra-thin-body FD SOI technology may be used to remove the need for circuit adaptation, thanks to its lower sensitivity against global process and temperature variations [10]. If the application features stand-by periods, sleep-mode power-gating technique should be considered for saving leakage energy. At all technology nodes, special care has to be taken when engineering the power switch for subthreshold operation [6].

Regarding future scaling, there are two possible scenarios depending on the evolution of the IC market. In the first scenario, the market growth (in volume) of low-performance applications remains slow. It is the authors' belief in this scenario that economical reasons will prevent from porting subthreshold circuits for ULP applications beyond 45nm node. Although further scaling will reduce the costs of raw material from die area reduction, the increasing costs associated to the manufacturing process and masks will no longer be supported by the niche market of low-performance ULP applications (low selling price of manufactured chips), when reaching 32/22 nm nodes. The second scenario is based on the assumption of the explosion of today's emerging market of ambient intelligence, as described in [11,12], for instance. In this case, this market would require billions of tiny ULP devices (sensor swarms), which might boost investments in 22nm technology node and below for those applications, which might then use subthreshold logic.

Mid-performance ULP applications

As shown in Fig. 1, in 130/90nm technology, performance issues prevent from operating at the minimum-energy point at f_{target} compatible with mid-performance applications. However, the minimum-energy point in 65/45nm high-performance/general-purpose (HP/GP) technologies is shifted towards higher f_{target} values [1], which might then enable subthreshold operation in mid-performance applications. This creates new opportunities for minimum-energy subthreshold circuits. In HP/GP flavor, the E_{min} level is higher at 65/45nm than at 130/90nm node [3]. An optimum MOSFET selection (low V_t and upsized gate length) allows to improves the key technological targets and fixes this problem [3]. It has thus to be used in nanometer subthreshold circuits for mid-performance applications. Adaptive RBB has to be used too for circuit adaptation even if the bias voltages may be larger in HP/GP than in LP flavor because of reduced body-bias coefficient [1]. Nevertheless, in mid-performance portable consumer applications, the form-factor and cost constraints are slightly relaxed as compared to low-performance applications. Generator or external supplies of larger negative bias voltages could thus be tolerated, which favors reverse ABB over forward ABB as it is more energy efficient [1]. Alternatively, an undoped-channel ultra-thin-body FD SOI technology may be used for further E_{min} improvement thanks to better subthreshold MOSFET characteristics. It also limits the need for circuit adaptation, which may lead to cost savings. Notice that depending on the I_0 value, near-threshold operation or architectural techniques such as parallelization or pipelining might be required for meeting the timing constraints.

To the authors' point of view, subthreshold circuits at future 32/22 nm nodes will not be feasible - or at least inefficient - in bulk technology. Indeed, bulk MOSFETs are likely to feature bad subthreshold characteristics at these nodes and bulk technology will no longer be able to meet the key technological targets for low E_{min}. Therefore, undoped-channel ultra-thin-body FD SOI technology will be compulsory for nanometer subthreshold circuits at 32/22nm. Moreover, as the market of portable consumer/professional applications is a mass production market, we think that it may motivate IC foundries to develop a process dedicated to subthreshold operation. This process should thus target optimum MOSFET characteristics to implement the specifications we reported in Section 2. At these nodes, it is also likely that circuit adaptation will be compulsory to avoid prohibitive design margins on V_{dd}, even in FD SOI technology. In order to provide adaptation opportunity, the technology should thus

come with an ultra-thin-buried-oxide (UTBOX) and dual back-gate (BG) bias [13], or with a double gate (DG) in independent-gate configuration [14]. Additionally, to the authors' point of view, multiple-gate devices such as FinFETs, MuGFETs or double-gate MOSFETs in common-gate configuration are less valuable for subthreshold logic for two reasons. First, back-gate biasing has a weak impact on these devices [6] and circuit adaptation through I_0 tuning can thus hardly be achieved. Second, the parasitic capacitances associated to the multiple gates [7] may increase the mean load capacitance C_L proportionally more than at nominal V_{dd} given the low instrinsic gate capacicance in subthreshold regime $C_{g,sub}$.

REFERENCES

[1] D. Bol, D. Flandre and J.-D. Legat, "Technology flavor selection and adaptive techniques for timing-constrained 45nm subthreshold circuits", in Proc. ACM/IEEE Int. Symp. Low-Power Electron. Des, 6 p., 2009.

[2] D. Bol, R. Ambroise, D. Flandre and J.-D. Legat, "Interests and limitations of technology scaling for subthreshold logic", in IEEE Trans. on VLSI Syst., in press, 12 p., 2009.

[3] D. Bol, D. Kamel, D. Flandre and J.-D. Legat, "Nanometer MOSFET effects on the minimulm-energy point of 45nm subthreshold logic", in Proc. ACM/IEEE Int. Symp. Low-Power Electron. Des, 6 p., 2009.

[4] D. Bol, R. Ambroise, D. Flandre and J.-D. Legat, "Sub-45nm fullydepleted SOI CMOS subthreshold logic for ultra-low-power applications", in Proc. IEEE International SOI Conference (Best Poster Award), pp. 57-58, 2008.

[5] D. Bol, R. Ambroise, D. Flandre and J.-D. Legat, "Analysis and minimization of practical energy in 45nm subthreshold logic circuits", in Proc. IEEE Int. Conf. Comp. Des. (Best Paper Award), pp 294-300, 2008.

[6] D. Bol, "Pushing ultra-low-power digital circuits into the nanometer era", Ph.D dissertation, Université catholique de Louvain, Dec. 2008.

[7] J. Srinivasan, S. V. Adve, P. Bose and J. A. Rivers, "The impact of technology scaling on lifetime reliability", in Proc. Int. Conf. Dependable Systems and Networks, pp. 177-188, 2004.

[8] B. Zhai, L. Nazhandali, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, D. Blaauw, T. Austin: "A 2.60pJ/inst subthreshold sensor processor for optimal energy efficiency", in Dig. Tech. Papers VLSI Circuits, pp. 154-155, 2006.

[9] Y. Pu, J. Pineda de Gyvez, H. Corporaal and Y. Ha, "An ultra-low-energy/frame multi-standard JPEG co-processor in 65nm CMOS with sub/near-threshold power supply", in IEEE Int. Solid-Stace Circ. Conf., pp. 146-147, 2009.

[10] O. Weber et al., "High immunity to threshold voltage variability in undoped ultra-thin FDSOI MOSFETs and its physical understanding", in Dig. IEEE Int. Electron Dev. Meeting, in press, 4 p., 2008.

[11] H. De man, "The changing world of semiconductors: impact on research and education", keynote talk 2008.

[12] J. Rabey, "A brand new wireless day", keynote talk at Asian-Pacific Des. Autom. Conf., 2008.

[13] R. Tsuchiya et al., "Controllable inverter delay and suppressing Vth fluctuation technology in Silicon on thin BOX featuring dual back-gate bias architecture", in Dig. IEEE Int. Electron Dev. Meeting, pp. 475-478, 2007.

[14] M. Masahara et al., "Demonstration, analysis and device design considerations for independent DG MOSFETs", in IEEE Trans. Electron Dev., vol. 52, no. 9, pp. 2046-2053, Sep. 2005.

[15] F. Dauge, J. Pretet, S. Cristoloveanu, A. Vandooren, L. Mathew, J. Jomaah, B.-Y. Nguyen, "Coupling effects and channels separation in FinFETs", in Solid-State Electronics, vol. 48, no. 4, pp. 535-542, April 2004.

[16] J.-P. Raskin, T. M. Chung, V. Kilchytska, D. Lederer and D. Flandre, "Analog/RF performance of multiple gate SOI devices: wideband simulations and characterization", in IEEE Trans. Electron Dev., vol. 53, no. 5, pp. 1088-1095, May 2006.