Citation	N. Reynders, W. Dehaene, (2012), Variation-Resilient Building Blocks for Ultra-Low-Energy Sub-Threshold Design IEEE Transactions on Circuits and Systems II: Express Briefs, 59 (12), 898-902.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1109/TCSII.2012.2231022
Journal homepage	http://tcas2.polito.it
Author contact	nele.reynders@esat.kuleuven.be + 32 (0)16 321104

(article begins on next page)

Variation-Resilient Building Blocks for Ultra-Low-Energy Sub-Threshold Design

Nele Reynders, *Student Member, IEEE*, and Wim Dehaene, *Senior Member, IEEE*

Abstract—This paper presents the design of variation-resilient ultra-low-voltage circuits functioning at MHz-speed. By careful design, robust digital circuits operating in the sub-threshold region are achieved. The paper discusses circuit techniques to obtain building blocks that are able to overcome the high sensitivity to variations and the decreased current ratios in sub-threshold while retaining MHz-performance. The building blocks are successfully implemented in 2 chips fabricated in 90 nm CMOS technology. Measurements show that the variation-resilient designs are fully functional at ultra-low supply voltages and obtain clock frequencies in the MHz-range and sub-pJ energy consumptions.

I. INTRODUCTION

THE energy consumption of digital circuits can drastically be reduced by lowering their supply voltage [1]. Operating circuits in the sub-threshold region is the extreme case of supply reduction [2], [3]. The main advantage of sub-threshold circuits is the minimization of the dynamic energy consumption. However, due to the decreased currents in this region, the speed of these circuits is limited. Therefore, sub-threshold circuits are mostly adequate for applications with very limited energy budget but less stringent speed performance requirements. Nonetheless, achieving speeds in the MHz-region is preferable to make sub-threshold circuits attractive for industrial applications, e.g. in hearing aids and body sensor networks.

Two main issues can compromise the functionality of circuits operating in the sub-threshold region. First, the ratio between the on-current I_{on} and the leakage current I_{off} decreases severely at lower supply voltages. More precisely, Fig. 1a shows the relevant I_{on}/I_{off} -ratios of LVT transistors as function of V_{dd} . The $I_{on,p}/I_{off,n}$ -ratio is always approximately a factor 100 smaller than the $I_{on,n}/I_{off,p}$ -ratio, and at low supply voltages it becomes problematically low. Second, sub-threshold circuits suffer from an exponential sensitivity to variations. This results in a deteriorated functionality and in highly variable gate delays. To illustrate the latter, the percentage variation under intra-die variations of the propagation delay of a regular sized CMOS inverter as function of V_{dd} obtained with Monte Carlo (MC) simulations is shown in Fig. 1b. An inverter that is sized for normal operation at a super-threshold supply displays a high increase in propagation delay variation when lowering V_{dd} , especially in the sub-threshold

Fig. 1. (a) Relevant I_{on}/I_{off} -ratios as function of the supply voltage. (b) Variation of propagation delay of a regular sized CMOS inverter as function of the supply, obtained with 1000 Monte Carlo simulations.

region. The classical way of dealing with variations is to take design margins to ensure yield. However, because of the high sensitivity to variations of sub-threshold circuits, accumulation of design margins compromises the low-power benefit of operating in sub-threshold. Therefore, this is not the optimal manner of coping with the increased variations. Recently, [4] has also proposed an alternative method that utilizes timing error-detection circuits in sub-threshold to account for both local and global variations.

Both issues can be overcome by careful circuit design, resulting in building blocks for a robust sub-threshold system. This paper discusses techniques for sub-threshold circuit design focusing on variation-resilience and on achieving ultra-low-energy at MHz-speed, while countering the decrease in current ratios. The paper compares different topologies used in state-of-the-art sub-threshold designs and discusses the various sub-threshold trade-offs. Furthermore, an in-depth analysis of the impact of variations is given. To validate the proposed techniques (i.e. transmission gate logic extended with nMOS stacking), they are implemented in two different designs [5], [6] of which measurement results are shared. These results serve as an illustration of the influence of different architectural design decisions. All results, both simulation and measurement, are obtained with a 90 nm CMOS technology.

Section II discusses in detail the circuit design of the sub-threshold building blocks. These building blocks were implemented in two sub-threshold datapath chips, of which section III describes the architectural choices and the measurement results. Finally, section IV concludes this paper.

II. DESIGN OF THE BUILDING BLOCKS

A. LVT transistors

Since modern deep sub-micron technologies offer multiple V_t -options, a choice has to be made concerning the threshold voltage selection. The inherent disadvantage of working in

The authors are with the MICAS division, Dept. of Electrical Engineering (ESAT), KU Leuven, 3001 Leuven, Belgium. W. Dehaene is also with IMEC, 3001 Leuven, Belgium (e-mail: nele.reynders@esat.kuleuven.be).

This work is supported by the Research Foundation - Flanders (FWO).

Copyright © 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

Fig. 2. Schematic of (a) a standard CMOS inverter sized for sub-threshold operation and (b) a stacked nMOS inverter with relaxed pMOS sizing.

Fig. 3. For $V_{dd} = 200$ mV: (a) Current percentage of stacked nMOS transistors relative to a single nMOS transistor, as function of the amount of stacked nMOS transistors. (b) Width of pMOS relative to nMOS (left axis) as function of the amount of stacked nMOS transistors in the CMOS inverter, in order to reach a maximal and balanced noise margin (right axis).

the sub-threshold region is the speed deterioration. However, by using low- V_t (LVT) transistors that have higher currents than standard- V_t (SVT) or high- V_t (HVT) transistors for the same supply voltage, a maximal sub-threshold speed can be guaranteed. Naturally, the use of LVT transistors also aggravates leakage. However, the increased leakage can be handled by taking this into account during circuit level design. The building blocks in this paper are therefore always constructed with LVT transistors, which is a first step to obtain sub-threshold circuits with MHz-range operating frequencies.

B. Inverter

Since the inverter is the most basic element, it is discussed as the first building block. To enable ultra-low-voltage operation for a standard CMOS inverter, the nMOS and pMOS transistor should be carefully balanced so that the noise margin is maximized [7]. Equalizing nMOS and pMOS strength is feasible by sizing the width of the pMOS transistor, often resulting in a very large relative width. In this 90 nm CMOS technology, the pMOS needs to be 11x wider than the nMOS in a standard CMOS inverter [5] (Fig. 2a). A solution to this issue is to employ nMOS stacking because stacking the nMOS transistor reduces both its I_{on} and its I_{off} . Fig. 3a shows which effect stacking has on the currents. As a result, the pMOS sizing can be relaxed without degrading the noise margin, as can be seen in Fig. 3b. The effect of stacking on the currents and thus on the pMOS sizing reduces with the amount of stacked transistors. Therefore, it is optimal to stack the nMOS transistor twice, resulting in a relative pMOS width of 6. Stacking not only allows to relax pMOS sizing, but also decreases the leakage through the nMOS transistor, which reduces the static power consumption. Using a stacked nMOS inverter (Fig. 2b) results in an increased nominal propagation delay of 23%, but this is outweighed by the total area reduction of 33% and the leakage power reduction of 61% compared to

Fig. 4. Variation of propagation delay of CMOS inverters with and without nMOS stacking as function of the supply, obtained with 1000 MC simulations.

the regular inverter (at $V_{dd} = 200$ mV). Moreover, the next paragraph discusses the positive effect that the stacked nMOS inverter has on delay variations.

Due to the exponential sensitivities to variations of sub-threshold circuits, the variation of the gate delay is an essential characteristic. It is shown that V_t -mismatch is the dominant source of delay variations in sub-threshold operation [8], [9]. Therefore, MC simulations were performed on the standard CMOS inverter versus the inverter with nMOS stacking (Fig. 4). Adequately sizing the standard CMOS inverter ($P=11$) to sub-threshold restrictions, clearly lowers the variation of propagation delay in the sub-threshold region compared to a regular sized inverter (see Fig. 1b). However, Fig. 4 also shows that using a stacked nMOS inverter further decreases the delay variation. To summarize, introducing nMOS stacking increases the nominal propagation delay slightly but it significantly reduces the variation of the delay. Due to the variation-resilience of the stacked nMOS inverter, lower design margins have to be introduced to cope with timing variations compared to conventional inverters. Therefore, the stacked nMOS inverter is the optimal choice of sub-threshold inverter topology.

C. Logic gates

The logic gate topology is a crucial choice for sub-threshold circuit design. A first option is to use standard CMOS logic. As previously mentioned, to acquire equal noise margins in the sub-threshold region, the pMOS width often needs to be sized excessively compared to the nMOS width. Logic gates that consist of stacked pMOS transistors (e.g. a NOR gate) hence require even more excessive pMOS sizing. Stacking the nMOS transistors of a NOR gate can relax the pMOS sizing, but the total area consumption is still high. Recently, [10], [11] have also suggested to increase the channel length to improve the transistor's sub-threshold behavior. However, balancing the pull-up and the pull-down transistors through length sizing to ameliorate sub-threshold functionality of standard CMOS logic is difficult. This is due to the high sensitivity of the current as function of the transistor's length to process and technology parameters. In general, transistor sizing at ultra-low-voltage operation is strongly technology dependent [12]. Additionally, the use of sub-threshold source-coupled logic [13] offers promising results, especially for nanometer-scale technologies, but this is beyond the scope of this paper.

Another option for sub-threshold logic gates proposed in [5] is the use of transmission gate (TG) logic extended with

Fig. 5. Schematic of (a) a standard CMOS NOR and (b) a TG NOR with nMOS stacking to reduce leakage.

nMOS stacking. As opposed to standard CMOS, there is no need for transistor balancing through sizing since there are always an nMOS and a pMOS included in a conducting path. To improve the problematically low $I_{on,p}/I_{off,n}$ -ratio and the imbalanced rise and fall times both nominally and under variations, the nMOS transistor is stacked. This results in a decreased leakage current $I_{off,n}$ [14] and thereby mitigates the current ratio problems. Without nMOS stacking, there is a large difference between the rise time and the fall time of a transmission gate because the minimal nMOS is much stronger than the minimal pMOS. Stacking the nMOS transistor of TG logic results in a more balanced rise and fall time and thus has no negative impact on the speed of the logic gate.

In the following analysis, standard CMOS logic with pMOS width upsizing (abbreviated to CMOS) and TG logic extended with nMOS stacking (abbreviated to TG) are compared on various logic gate characteristics. The analysis is performed on a NOR gate (Fig. 5) because it is an elementary logic function and a difficult gate in standard CMOS logic because it requires pMOS stacking. Since all logic gates in TG logic have the same generic structure [6], the results for other TG logic gates will be very similar to the ones of the NOR. In the analysis, the NOR is subjected to inter- and intra-die variations. Due to the exponential sensitivity to variations, it is of the utmost importance to design variation-resilient sub-threshold circuits.

Because of the small supply voltage swing, an important characteristic in sub-threshold design is the output signal loss of logic gates. Too much signal loss can cause the subsequent gate to wrongly interpret the logic value. Signal losses can be overcome by regenerating the signal, e.g. through an inverter. For example in a datapath with a high logic depth, intermediate signals of cascaded logic gates can be regenerated to ensure correct output levels. However, the lower the amount of signal loss, the less frequent inverters need to be inserted to restore the signal levels to the supply rails.

Fig. 6 compares a TG and a CMOS NOR gate on the percentage signal loss their output has relative to the total supply swing, under inter-die variations. Only the worst-case corners are shown as a function of V_{dd} . In the case of signal loss on logic low level, the logic gates perform worst in the slow-nMOS fast-pMOS (snfp) corner because of the weakened nMOS transistor versus the strengthened pMOS transistor. Respectively, at signal loss on logic high level, this worst-case applies to the fast-nMOS slow-pMOS (fnsp) corner. Fig. 6 shows that the signal loss aggravates when the supply voltage lowers and the circuits operate more in sub-threshold. It is clear that the TG NOR outperforms the CMOS NOR in signal loss on logic low level, and TG logic is also the better option

Fig. 6. Percentage signal loss for different topologies of a NOR gate in the worst-case corner of (a) logic low and (b) logic high level as function of V_{dd} .

Fig. 7. Cumulative Distribution Function of the signal loss for different topologies of a NOR of (a) logic low and (b) logic high level for $V_{dd} = 200$ mV, obtained with MC simulations around the typical-typical corner.

Fig. 8. Variation of propagation delay of different topologies of a NOR gate as function of the supply, obtained with 1000 MC simulations.

in the case of signal loss on logic high level. The output swing degradation analysis is also performed for intra-die variations by carrying out extensive MC simulations for a supply of 200 mV. Fig. 7 demonstrates that the TG NOR performs significantly better under intra-die variations for signal loss on logic low level and comparably for logic high level.

Another essential characteristic is the variation of sub-threshold gate delay. As previously mentioned, intra-die variations are the dominant deteriorating influence on the variation in delay. Therefore, Fig. 8 shows the variation of the propagation delay as function of V_{dd} . The TG NOR displays overall less delay variations than the CMOS NOR.

One of the other advantages of TG logic is that it uses considerably smaller transistor dimensions compared to standard CMOS logic while achieving better variation-resilience. Moreover, upsizing is often necessary to reduce the sensitivity to variations of ultra-low-voltage standard CMOS logic [15], [16]. These extra margins are not necessary for TG logic. An

Fig. 9. Schematic of cascade of n logic gates, followed by an inverter.

extra benefit of TG logic is that it does not have direct leakage paths from V_{dd} to the ground, as such a logic gate has almost no contribution to the total leakage power of a system. To conclude, TG logic is the most attractive solution for sub-threshold logic gates, taking into account variability. Consequently, TG logic is the building block for all logic gates.

D. Cascading logic gates

To obtain an area- and energy-efficient design, it is beneficial to cascade logic gates. However, Fig. 6 and 7 showed that signal losses are present in sub-threshold logic gates. As a result, by cascading too much logic gates, the robustness can be deteriorated because of too large output signal losses. As stated before, it is thus necessary to regenerate intermediate signal levels. The previously discussed stacked nMOS inverter can serve as such a regeneration circuit, as well as memory elements like latches or flip-flops. The TG logic consumes considerably smaller leakage power and dynamic energy than the regenerating elements. Therefore, maximizing the number of cascaded gates while guaranteeing functionality is beneficial in terms of energy consumption. Fig. 9 shows the simulation setup used to quantify how many logic gates can be cascaded without compromising functionality. The stacked nMOS inverter is used as regenerating element. Fig. 10 gives the simulation results for logic depths from 1 to 4 under intra-die variations. Since TG logic gates suffer significantly more from output losses on logic high level (as shown in Fig. 7), the figures only display the most pessimistic case where the output of the TG logic gates is high.

Signal losses do not necessarily pose a threat for functionality, as long as the inverter is able to interpret the logic level correct. Therefore, the spread of the input voltage level of the inverter is compared to the spread of the switching point of the inverter. In the case where the input should be a logic '1' but the input voltage of the inverter is lower than its switching voltage, the input is wrongly propagated. To evaluate the chance of this worst-case scenario, a criterion with a yield of 1 wrong propagation out of a billion propagations has been used. If for n logic gates, the chance on wrong propagation is smaller than 1 out of a billion, cascading n gates is considered to not compromise robustness. This yield $Y_{1/1\text{billion}}$ is equal to $(1 - \text{normcdf}(6))$ for a single normal distribution, thereby corresponding to 6σ . But, for two uncorrelated distributions, the yield $Y_{1/1\text{billion}}$ is equal to $(1 - \text{normcdf}(4))^2$, hence corresponding to 4σ for each distribution. Therefore, both distributions of the output level of the cascade and the switching point of the inverter are evaluated at 4σ : the red vertical lines on Fig. 10 show the values of $\mu - 4\sigma$ of the input voltage and $\mu + 4\sigma$ of the switching voltage. If the two values coincide, a yield of 1 out of a billion wrong propagations is reached.

Fig. 10. Simulation results from the test setup in Fig. 9: Cumulative Distribution Function of the output voltage level (IN) of a cascade of n logic gates, compared to the switching point V_m of the inverter and the inverter's output (OUT), obtained with extensive MC simulations for $V_{dd} = 150$ mV.

To make sure the cascade of logic gates functions under all circumstances, the simulation is carried out for the target minimal supply voltage of the building blocks, i.e. 150 mV. Fig. 10 demonstrates that for logic depths of 1 and 2 there is still margin on the $Y_{1/1\text{billion}}$ -criterion because $\mu_{IN} - 4\sigma_{IN} > \mu_{V_m} + 4\sigma_{V_m}$. Consequently, cascades of 1 or 2 logic gates have a chance on wrong propagation that is smaller than 1 out of a billion. Fig. 10 also shows that for cascades of 3 and 4 logic gates $\mu_{IN} - 4\sigma_{IN} < \mu_{V_m} + 4\sigma_{V_m}$. Therefore, cascading more than 2 logic gates results in a deteriorated yield. It can also be seen that the output of an inverter after a cascade of 4 logic gates displays a higher spread than the spread of its input, thereby making it highly discouraging to use such logic depth. To conclude, for a 150 mV supply, the criterion indicates that for a cascade of 2 logic gates the trade-off between energy consumption and guaranteed robustness is optimal. However, for a higher supply voltage, the maximum logic depth to guarantee reliable operation also increases. A reduction of the number of regenerating elements will thus be obtained by redoing the analysis with a higher target supply voltage (e.g. 200 mV) to increase the maximum logic depth.

III. SUB-THRESHOLD DESIGNS

To test the building blocks, two designs of datapath elements were fabricated in the same 90 nm CMOS technology.

A. Design 1

The first design consists of a 32 bit logarithmic adder [5]. The adder is constructed with the building blocks described in this paper, in a pipelined architecture. Each pipeline stage is composed of 1 TG logic gate. Latch-based pipelining is employed because such an architecture allows time borrowing, i.e. the ability of slower logic using calculation time of faster logic [17]. When some logic blocks require a longer

Fig. 11. Measurement results of multiple dies of the two designs: average clock frequency and energy consumption as function of V_{dd} .

calculation time than others, a latch-based pipeline will tend to operate at the average of the delays. Since sub-threshold logic gates suffer from highly variable gate delays, time borrowing partly mitigates this variability problem.

Measurement results of multiple dies [5] show that the building blocks have been successfully validated in the sub-threshold adder. The adder is fully functional down to a supply voltage of 190 mV, thereby confirming that the building blocks are operational in the sub-threshold region, as well as variation-resilient. At the minimal V_{dd} , the clock frequency is 10 MHz and the energy consumption per addition is 0.4 pJ, resulting in an Energy-Delay Product (EDP) of 0.043 pJ. μ s (Fig. 11). These results demonstrate that it is possible to achieve MHz-speed while obtaining ultra-low energy consumption.

B. Design 2

The second design is a 16 bit Multiply-Accumulate (MAC) [6], which is a more complex datapath element that contains feedback. A new architectural strategy with the same building blocks is used to further improve performance and variation-resilience. The pipeline stage length is increased by cascading multiple TG logic gates, thereby decreasing the number of latches. Therefore, all TG logic gates are implemented differentially which enhances the variation-resilience and eases the design of the latch. Most importantly, an extra measure is obtained to reduce the effect of timing variations in the sub-threshold region. [8] showed that by increasing logic depth, the effects of individual gate variations on total path delay and energy reduces. The MAC implements this averaging of variations by cascading 2 differential TGs between the latches.

Improved measured results of multiple dies [6] are obtained by using the new strategy. The MAC is functional down to a significantly lower V_{dd} of 150 mV compared to the adder due to the enhanced variation-resilience. The minimal energy consumption per operation of 0.87 pJ occurs at a 190 mV supply, at an EDP of 0.083 pJ. μ s. A frequency comparison between the designs shows that at the same V_{dd} of 190 mV, the adder and the MAC are both able to operate at a clock of 10 MHz (Fig. 11). The impact of timing variations is thus drastically reduced, because the clock frequency remains equal although the pipeline stage length is doubled in the MAC.

IV. CONCLUSION

This paper presented the design of digital circuits operating in the sub-threshold region. The circuits were targeted to achieve an ultra-low sub-pJ energy consumption combined with MHz-speed. Techniques for the circuit design of building blocks have been thoroughly discussed, focusing on variation-resilience and robustness. The building blocks have been validated in two designs fabricated in a 90 nm CMOS technology. Both designs were fully functional at ultra-low supply voltages, working at clock frequencies in the MHz-region and achieving very low EDP figures. Architectural choices made it possible to increase variation-resilience by implementing time borrowing and averaging of timing variations.

ACKNOWLEDGMENT

The authors would like to thank Stefan Cosemans and Bram Rooseleer for their valuable technical input.

REFERENCES

- [1] D. Markovic *et al.*, "Ultralow-Power Design in Near-Threshold Region," *Proceedings of the IEEE*, vol. 98, no. 2, pp. 237–252, Feb. 2010.
- [2] H. Soeleman and K. Roy, "Ultra-Low Power Digital Subthreshold Logic Circuits," in *Proc. Int. Symp. on Low Power Electronics and Design (ISLPED)*, Aug. 1999, pp. 94–96.
- [3] A. Wang, B. Calhoun, and A. Chandrakasan, *Sub-threshold Design for Ultra Low-Power Systems*. Springer, 2006.
- [4] J. Mäkipää *et al.*, "Timing-Error Detection Design Considerations in Subthreshold: An 8-bit Microprocessor in 65nm CMOS," *J. Low Power Electronics and Applications*, vol. 2, no. 2, pp. 180–196, 2012.
- [5] N. Reynnders and W. Dehaene, "A 190mV supply, 10MHz, 90nm CMOS, Pipelined Sub-Threshold Adder using Variation-Resilient Circuit Techniques," in *Proc. Asian Solid-State Circuits Conference (A-SSCC)*, Nov. 2011, pp. 113–116.
- [6] N. Reynnders and W. Dehaene, "Variation-Resilient Sub-Threshold Circuit Solutions for Ultra-Low-Power Digital Signal Processors with 10MHz Clock Frequency," in *Proc. European Solid-State Circuits Conference (ESSCIRC)*, Sep. 2012, accepted for publication.
- [7] M. Alioto, "Understanding DC Behavior of Subthreshold CMOS Logic Through Closed-Form Analysis," *IEEE Trans. on Circuits and Systems I*, vol. 57, no. 7, pp. 1597–1607, Jul. 2010.
- [8] B. Zhai *et al.*, "Analysis and Mitigation of Variability in Subthreshold Design," in *Proc. Int. Symp. on Low Power Electronics and Design (ISLPED)*, Aug. 2005, pp. 20–25.
- [9] F. Frustaci, P. Corsonello, and S. Perri, "Analytical Delay Model Considering Variability Effects in Subthreshold Domain," *IEEE Trans. on Circuits and Systems II*, vol. 59, no. 3, pp. 168–172, Mar. 2012.
- [10] D. Bol *et al.*, "Interests and Limitations of Technology Scaling for Subthreshold Logic," *IEEE Trans. on Very Large Scale Integration Systems*, vol. 17, no. 10, pp. 1508–1519, Oct. 2009.
- [11] S. Hanson *et al.*, "Nanometer Device Scaling in Subthreshold Logic and SRAM," *IEEE Trans. on Electron Devices*, vol. 55, no. 1, pp. 175–185, Jan. 2008.
- [12] M. Alioto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: A Tutorial," *IEEE Trans. on Circuits and Systems I*, vol. 59, no. 1, pp. 3–29, Jan. 2012.
- [13] A. Tajalli and Y. Leblebici, "Leakage Current Reduction Using Sub-threshold Source-Coupled Logic," *IEEE Trans. on Circuits and Systems II*, vol. 56, no. 5, pp. 374–378, May 2009.
- [14] S. Narendra *et al.*, "Scaling of Stack Effect and its Application for Leakage Reduction," in *Proc. Int. Symp. on Low Power Electronics and Design (ISLPED)*, Aug. 2001, pp. 195–200.
- [15] J. Kwong and A. Chandrakasan, "Variation-Driven Device Sizing for Minimum Energy Sub-threshold Circuits," in *Proc. Int. Symp. on Low Power Electronics and Design (ISLPED)*, Oct. 2006, pp. 8–13.
- [16] Y. Pu *et al.*, "Vt Balancing and Device Sizing Towards High Yield of Sub-threshold Static Logic Gates," in *Proc. Int. Symp. on Low Power Electronics and Design (ISLPED)*, Aug. 2007, pp. 355–358.
- [17] D. Harris, *Skew-Tolerant Circuit Design*. Morgan Kaufmann Pub., 2001.