

Exploration and Design of Low-Energy Logic Cells for 1 kHz Always-on Systems

Maxime Feyerick, Jaro De Roose, Marian Verhelst
 ESAT-MICAS, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven (Heverlee), Belgium
 Email: maxime.feyerick@esat.kuleuven.be

Abstract—A standard cell library targeting always-on operation at 1 kHz is designed at circuit-level. This paper proposes a design methodology to achieve robust operation with minimum energy. Such minimum energy per operation for always-on systems is achieved by one specific supply and threshold voltage V_{Th} combination. As V_{Th} is discrete in a practical bulk technology, this minimum can however not be achieved through simple voltage tuning. In the considered 90 nm CMOS technology, V_{Th} is too low resulting in leakage dominated systems and preventing from attaining the minimum energy point in subthreshold. Three circuit techniques are optimally combined to fight leakage: stacking, reverse body biasing and optimal transistor dimensioning relying on second order effects of the dimensions on V_{Th} . They jointly allow logic gates to achieve the best balance between dynamic and leakage power. Moreover, the paper presents modified flip-flop topologies that also reliably operate at 0.27 V along with the gates. Benefits of improved logic gates and flip-flops are demonstrated on a small always-on feature-extraction system calculating running average and variance on a 1 Ksample/s data stream. The resulting system consumes 162 pW in simulation, or two orders of magnitude less when compared to a commercial library at its 1 V nominal voltage, or 1 order of magnitude less when compared to the commercial library at the same 0.27 V operating voltage.

I. INTRODUCTION

The recent trend of Internet of Things (IoT) devices has unveiled new application domains of low power continuous monitoring sensors. Recent examples of such devices are always-on acoustic object recognition [1], wireless sensor networks [2], continuous measuring by biomedical implants [3], etc. The small form factor of such devices severely limits the capacity of batteries included in the package. Alternatively, energy scavenging is used, which offers only a power source in the order of nanowatts. On the other hand, these always-on devices require generally a relatively low throughput and operate in the kHz-range.

Unfortunately, using commercially available low-power Standard Cell Libraries (SCL) for kHz-range systems results in severely leakage dominated designs in recent technology nodes. The faster cells spent a majority of the time leaking while waiting for the relatively low clock period. Duty-cycling the hardware by e.g. power-gating is often not possible because of the leakage power of memory elements needed to retain the state of the system between subsequent samples.

This work presents a design methodology optimizing energy consumption of combinational and sequential logic cells operating in always-on at 1 kHz. This paper is organized as follows. First Section II theoretically analyses the energy

consumption in always-on system. Section III addresses circuit techniques necessary to remedy leakage. Further, in Section IV, these techniques are used to design gates with balanced dynamic and leakage power. Section V considers robust flip-flops to accompany the gates. Finally, the designed cells are deployed in a practical sensing system to benchmark them against a commercial library in Section VI.

II. ENERGY IN ALWAYS-ON CMOS SYSTEMS

A. Theory

The energy per operation E_{op} of a gate consists out of two sources: the dynamic energy E_{dyn} and the leakage energy E_{leak} .

$$E_{dyn} = \alpha C_{gate} V_{dd}^2 \quad (1)$$

$$E_{leak} = P_{leak} / f_{clk} = I_{off} V_{dd} / f_{clk}, \quad (2)$$

where α is the activity, f_{clk} the clock frequency and C_{gate} the parasitic capacitance of the gate. E.g. for a minimum size inverter in 90nm CMOS technology $C_{inv} = 0.76$ fJ. I_{off} is the average leakage current of the gate. The maximal clock frequency f_{max} is determined by the logic depth d_L of the critical path and the (V_{dd}, V_{Th}) operating point:

$$f_{max} \approx \frac{I_{on}(V_{dd}, V_{Th})}{d_L C_{inv} V_{dd}}. \quad (3)$$

In always-on sensing systems, the interaction with the environment dictates that f_{clk} should equal a fixed frequency f_{req} , not necessarily equal to the maximum frequency f_{max} , with a target f_{req} of 1 kHz in this paper. Duty-cycled or high-performance systems, on the contrary, do not have this constraint, and can freely set $f_{clk} = f_{max}$. Having $f_{clk} < f_{max}$ is non-optimal regarding energy consumption: f_{max} can be lowered by e.g. voltage scaling until $f_{clk} = f_{max}$. As such, in always-on systems, both constraints should be fulfilled for minimum energy operation: $f_{clk} = f_{max} = f_{req}$. The energy minimization problem can thus be described as:

$$\begin{aligned} & \min_{V_{th}, V_{dd}} \left(E_{dyn} + \frac{P_{leak}}{f_{max}} \right) \\ & \text{subject to } f_{max}(V_{dd}, V_{Th}) = f_{req}. \end{aligned} \quad (4)$$

Filling in (1) and (2) in the objective function in (4) removes its dependence on V_{Th} , assuming I_{on} is a subthres-

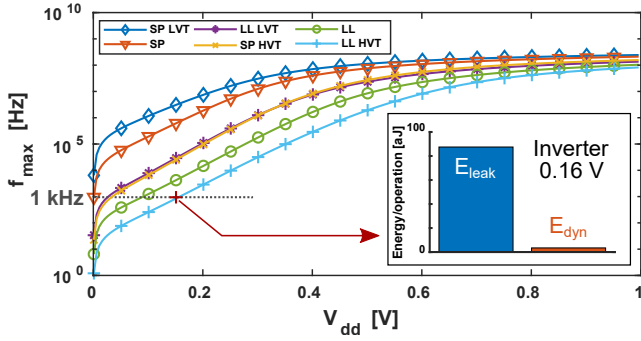


Fig. 1. A: Maximal clock frequency for every flavor as a function of supply voltage. B: Energy composition for an unsized inverter at 0.16 V using LL-HVT transistors. $\alpha = 0.2$, $d_L = 250$.

hold current with $I_{on} = I_0 \frac{W}{L} e^{(V_{dd}-V_{Th})/(nv_t)}$ and $I_{off} = I_0 \frac{W}{L} e^{-V_{Th}/(nv_t)}$:

$$\min_{V_{th}, V_{dd}} \left(\alpha C_{inv} V_{dd}^2 + e^{\frac{-V_{dd}}{nv_t}} C_{inv} V_{dd}^2 d_L \right) \quad (5)$$

$$\text{subject to } \frac{I_0 \frac{W}{L} e^{\frac{V_{dd}-V_{Th}}{nv_t}}}{d_L \cdot C_{inv} \cdot V_{dd}} = f_{req}. \quad (6)$$

The solution to (5) is the same Minimum Energy Point (MEP) as known in systems with unconstrained frequency [4]. However, in always-on systems, (6) instructs that V_{Th} is not free to choose if one wants to operate the system at a target speed at the MEP. Given a frequency f_{req} , and an architecture described by α and d_L , only one (V_{dd}, V_{Th}) pair achieves minimal energy consumption.

B. Practical Technological Limitation and Consequences

V_{Th} is treated as a continuous parameter in the above derivation. This is in sharp contrast to every practical technology where V_{Th} is a discrete parameter offered under the form of different so-called technology flavors. Two classes of flavors exist in the targeted 90 nm process, Standard Process (SP) and Low Leakage (LL), the latter having higher V_{Th} . Each class itself is further subdivided in three flavors. Fig. 1 shows the maximal frequency in function of V_{dd} for different flavors. To calculate f_{max} , typical values for the architectural parameters α and d_L are assumed.

Of all flavors, LL-HVT needs the highest supply to have $f_{max} = 1$ kHz, i.e. $V_{dd} \approx 0.16$ V. This is deep in subthreshold which justifies the assumptions made in (5)-(6). Fig. 1b shows for an unsized inverter that under these conditions leakage still dominates at this voltage, indicating the MEP is at a higher V_{dd} . Yet, V_{dd} cannot simply be raised as this would violate constraint (6) without also raising V_{Th} , resulting in even higher energy consumption at $f_{clk} = f_{req}$. This explains the leakage dominance of typical SCLs.

III. CIRCUIT TECHNIQUES TO REDUCE LEAKAGE

A. The Effect of Circuit Techniques

Various circuit techniques are selected to reduce I_{off} and help balance E_{leak} and E_{dyn} . However, many also negatively

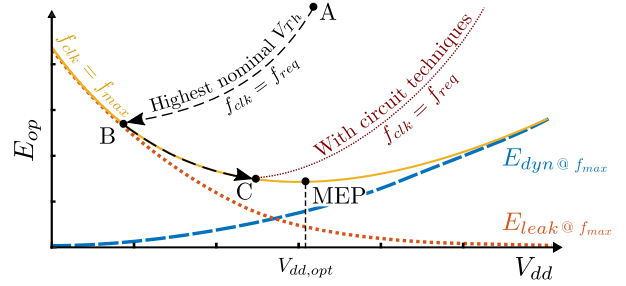


Fig. 2. Effect of circuit techniques on an always-on system with too low V_{Th} for operating at 1 kHz. Point A is with too low V_{Th} and too high V_{dd} , resulting in $f_{max} > f_{req}$. By voltage scaling, B is reached where $f_{max} = f_{req}$. By applying a technique, C can be reached, also with $f_{max} = f_{req}$, yet much nearer to the MEP.

impact I_{on} to the same degree as I_{off} , as its conduction mechanism is identical in subthreshold. Therefore, the energy in the MEP does not improve because $E_{leak} \propto I_{off}/I_{on}$. It even rises moderately as most techniques introduce a penalty on C_{inv} .

Still, the techniques are useful in providing a substitute for the further required raise in V_{Th} , making the MEP reachable under the constraint $f_{clk} = f_{req}$. Figure 2 shows how a technique can slow down the system, shifting the curve where $f_{clk} = f_{req}$ to higher V_{dd} . Consequently, when voltage scaling, the point C where $f_{clk} = f_{max}$ lies now closer to the MEP.

B. Employed Techniques

1) *Optimal Transistor Dimensions*: Nanometer-CMOS introduces a set of second-order effects of the transistor dimensions on V_{Th} . The Reverse Short Channel Effect (RSCE) causes a rise of V_{Th} for diminishing transistor length L . It is a consequence of halo-doping [5]. It interacts with the Short Channel Effect (SCE), which on the contrary lowers V_{Th} for diminishing L [6]. The combination gives a characteristic dependence of V_{Th} on L , where with decreasing L , V_{Th} first rises and eventually falls. A peak in V_{Th} corresponds with a minimum in I_{off} .

Similarly, the Reverse Narrow Channel Effect (RNCE) causes a decrease of V_{Th} for decreasing transistor width W [7]. In combination with the usual W -factor in I_{off} and I_{on} again a minimum can arise, or at least a diminished increase of I_{on} and I_{off} for small W . Another consequence is that in subthreshold, W is not a very effective knob to increase drive strength [8]. Therefore, this work uses transistor fingers instead of W to increase the drive strength of gates.

These effects are technology and flavor dependent. Figure 3 shows I_{off} for the LL-HVT NMOS as a function of both W and L in the targeted 90 nm technology. Taking the effects jointly into account, the optimal dimensions can be found by minimizing E_{op} at $f_{clk} = f_{req} = 1$ kHz. This is indicated by the red plus sign. These dimensions are smaller than the minimum I_{off} (red circle) due to the reduced parasitic capacitance. The corresponding dimensions are as follows: $W_{NMOS} = 216$ nm, $L_{NMOS} = 153$ nm, $W_{PMOS} = 120$ nm and $L_{PMOS} = 405$ nm.

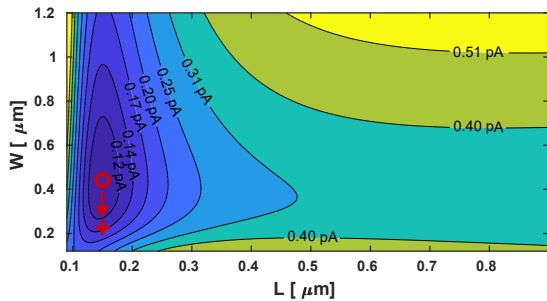


Fig. 3. Dependence of I_{off} (LL-HVT NMOS) on channel length and width. The circle indicates the minimum, the plus sign the point where E_{op} is minimal at 1 kHz. $V_{dd} = 0.27$ V.

2) *Stacking*: Stacking is a well-known and effective leakage-reducing technique in superthreshold. Forced stacking is considered, where one transistor is replaced by two in series. The induced voltage on the internal node connecting drain and source of both transistors causes the staking effect [9]. In superthreshold circuits, the stacking effect affects I_{off} more than the superthreshold current I_{on} because of the exponential dependence on the voltage of the internal node for subthreshold currents. In subthreshold on the other hand, the effect on I_{on} and I_{off} is almost identical and weaker than in superthreshold [8]. Still, the technique is useful to slow down the system such that $f_{max} = f_{req}$ in the MEP.

3) *Body Biasing*: Applying a Reverse Body Bias (RBB) to the bulk of a MOSFET enhances V_{Th} to make up for the lower V_{Th} . A few disadvantages exist. Firstly, the sensitivity of V_{Th} for a bulk-source bias decreases with newer nanometer technology nodes [10]. RBB also requires charge pumps to generate the body voltage above and beneath the supply rails and two nets to distribute the voltages. Finally, a triple well is needed to freely choose the NMOS bulk voltage, incurring an area overhead. On the other hand, the subthreshold currents are sensitive for small variations in V_{Th} due to the exponential dependency. Simulations show with $|V_{BS}| = 0.5$ V a $\Delta V_{Th} \approx 0.70$ mV for both NMOS and PMOS LL-HVT transistor, resulting in respectively $4.5\times$ and $12.3\times$ less I_{off} .

IV. METHODOLOGY FOR GATE DESIGN IN ALWAYS-ON SYSTEMS

A. Design Requirements

Operating in subthreshold is challenging for reliability reasons [11], [8]. The Noise Margin NM is used as a measure for correct functioning of a gate. A small NM indicates low tolerance for static noise on the input signals of a gate. In subthreshold specifically, the noise margins are compromised due to low swing and high sensitivity for process variations. Besides the low supply, the imbalance between PUN and PDN is another important factor reducing NM [8]. Fig. 4 indicates that, in nominal conditions, a large imbalance between PMOS and NMOS exists in the target technology. For an inverter, the NMOS should be sized $8\times$ larger at the approximate target supply to undo the imbalance, implying an area and parasitic

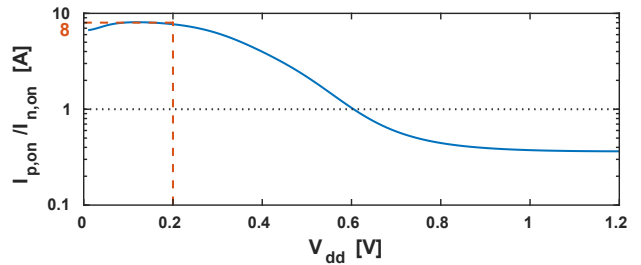


Fig. 4. Ratio of I_{on} of minimum sized PMOS over NMOS in considered 90 nm technology (logarithmic y-scale).

capacitance cost. The worst-case imbalance grows further under statistical variations, impairing possible yield [8]. As from the perspective of an SCL, the number of cells and their sizes in a design, and the needed margins to ensure a specific yield under statistical variations are application-specific and thus unknown, it is chosen here to maximize the noise margin. This way, yield is ensured under all circumstances. Maximizing NM is done by upsizing the PDN though transistor fingers to avoid RNCE, until the gate is balanced, i.e. $NM_L = NM_H$.

Besides static robustness, also dynamic performance must be guaranteed. Gates are designed ensuring that the required $f_{clk} = 1$ kHz is met at the specified supply voltage. For the different inverter designs, the supply is chosen as to just meet the speed requirement of the supposed architecture ($d_L = 250$) in the worst-case process corner (i.e. the ss-corner). The gate delay is simulated using a capacitive load of four inverters plus the internal load as to provide a realistic load, equaling the FO4-delay [12] in case of an inverter. As such, for an inverter, V_{dd} is determined by:

$$f_{clk} = \frac{4 + 1}{d_L t_{FO4,inv}} = 1 \text{ kHz} \Leftrightarrow t_{FO4,inv}(V_{dd}) = 20 \mu\text{s}. \quad (7)$$

The supply fulfilling (7) is used to evaluate the E_{dyn} and E_{leak} for each inverter design.

B. Comparison of Techniques on an Inverter

The individual techniques are compared for a balanced inverter in fig. 5. Stacking and unequal body bias are only applied on the PUN as the PMOS is the strongest according to fig. 4. This helps to reduce the fingers in the PDN while achieving balancing. The body bias voltages are for NMOS and PMOS respectively -0.4 V and $V_{dd} + 0.4$ V in case of equal body biasing, or 0 V and $V_{dd} + 0.4$ V in case of unequal body biasing.

From fig. 5, it is clear that E_{op} is drastically improved by the techniques as they reduce the dominant leakage contribution. With some techniques the dynamic energy also decreases despite the supply rise, due to the reduced number of fingers needed in the PDN to balance the gate. The supply voltage and necessary fingers in the PUN are summarized in table I.

Unequal body bias is the most effective technique, as RBB solely on the PMOS is used to completely remove the need for balancing through fingers. Optimal transistor dimensions are more effective than stacking but come at a higher price in

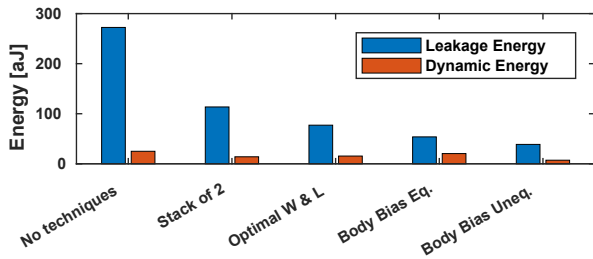


Fig. 5. Comparison of E_{dyn} and E_{leak} in inverter designs with single techniques applied. $\alpha = 0.2$ for all gate E_{dyn} -figures.

TABLE I
FINGERS IN THE PDN AND SUPPLY VOLTAGE IN BALANCED INVERTERS USING A SINGLE TECHNIQUE.

	Number of PMOS Fingers	V_{dd}
No technique	8	0.215 V
Stack of 2	3	0.225 V
Optimal W & L	3	0.205 V
Equal Body Bias	4	0.27 V
Unequal Body Bias	1	0.235 V

area. Overall, no individual technique is sufficient to achieve balance between leakage and dynamic energy. Techniques are combined in the final design, which is discussed in next section.

C. Optimally Combining Techniques

1) *Inverter*: Fig. 6 compares different inverter designs that combine the previously discussed techniques. Only one finger is necessary to balance each inverter as the combined techniques primarily target the PUN. V_{dd} varies from 0.225 V to 0.275 V to meet the timing constraint when progressively applying techniques. The bulk voltages of each inverter are adapted to achieve balanced NM . None of the combinations of two techniques is sufficient to adequately balance leakage and dynamic energy either. The combination of a stack of 3 and optimal dimensions shows that increasing stacking further has diminishing returns. Combining the three techniques finally balances leakage and dynamic energy. The combination of the three techniques with smaller adapted transistor dimensions reduces area and parasitic capacitance by applying more body bias, allowing a further reduction of both E_{dyn} and E_{leak} . The new transistor dimensions are as follows: $W_{NMOS} = 180$ nm, $L_{NMOS} = 153$ nm, $W_{PMOS} = 120$ nm and $L_{PMOS} = 180$ nm. This configuration is the preferred option, and choosing it fixes the supply on 0.27 V. The area is three times that of an inverter balanced for superthreshold in the same technology. Table II summarizes among other gates the results of the final inverter design in typical conditions. Under process corners, the worst-case delay equals 18.9 μ s in the ss-corner. The worst-case noise margins are $NM_L = 65.70$ mV and $NM_H = 67.25$ mV in respectively the fnsp- and snfp-corners.

2) *Complex Gates*: NOR- and NAND-gates are also designed. Together with the inverter and the flip-flop designed in section V, they form a minimal but complete SCL. The

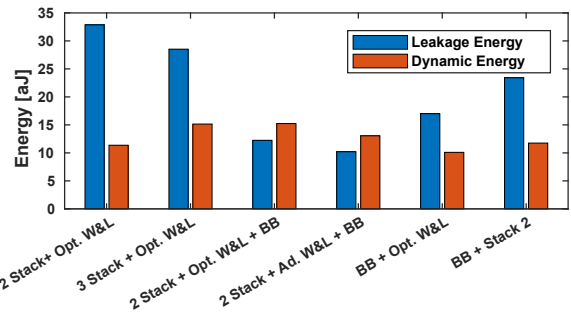


Fig. 6. Comparison of E_{dyn} and E_{leak} in inverters combining techniques.

TABLE II
SPECIFICATIONS OF PROPOSED GATES, TYPICAL CONDITIONS, USING STACKING, ADAPTED DIMENSIONS AND BODY BIAS. FO4-LOAD, 0.27 V.

	$t_{d,max}^*$ [μ s]	$E_{leak,avg}$ [aJ]	E_{dyn} [aJ]	$E_{tot,1}$ kHz [aJ]	NM_L [mV]	NM_H [mV]	Area [μ m ²] [†]
INV	2.578	10.21	13.05	23.26	100.4	101.8	3.275
NAND	4.370	10.96	16.95	27.91	100.6	100.9	6.551
NOR	4.635	11.08	20.45	31.53	99.58	101.9	6.551

* The average of the maximum rise and fall delay.

[†] Relative to a minimum size inverter balanced for superthreshold (216 000 nm²).

conclusions are very similar to the inverter. A combination of all three techniques is again necessary to balance E_{dyn} and E_{leak} . The same body bias and supply voltage are used as with the inverter. The specifications are given in table II.

V. FLIP-FLOP DESIGN

A. Considered Topologies and Adaptations for Subthreshold

A selection of master-slave flip-flops and pulse-triggered latches is modified and compared to find the optimal topology for subthreshold. The selection of master-slave latches consists of the PowerPc [13], C²MOS and Transmission Gate flip-flops [14]. As pulse-triggered latches, the Hybrid Latch Flip-Flop (HLFF) [15] and Sense-Amplifier Flip-Flop (SAFF) [16] are considered.

However, subthreshold operation requires modifications to the topologies due to the strong unbalance of NMOS and PMOS in subthreshold and high sensitivity for V_{Th} -variations. Fig. 7 shows the different topologies with the here proposed modifications. ‘LL’ indicates that the transistor is replaced by an LL(-RVT)-flavored one, which has a lower V_{Th} than the LL-HVT. MOSFETs upsized by using fingers are accompanied by their scale factor. The master-slave latches are modified by reinforcing the PDN of some internal (tri-state) inverters. The stacked LL NMOS offer a drive strength which is balanced with the PMOS over a large supply interval from 0.1 V up to 1 V. The PDN is reinforced instead of weakening the PUN to not compromise speed too much. Inverters at the input are balanced to ensure balanced t_{setup} for both high-low and low-high transitions. Inverters driving the output load under certain conditions are also balanced. The main disadvantage of LL transistors is enhanced leakage.

The HLFF employs contention as a way for the pulse-triggered stage to write the cross-coupled inverter at the output

in the short transparency window of the flip-flop [15]. The output of the pulse-triggered stage must overpower the weak inverter of the pair in this short window. To facilitate this, two measures are taken: 1) the window is enlarged by adding a PMOS pass gate in the Clkb generation, 2) the output of the pulse-triggered stage is carefully balanced with the weakly sized inverter of the cross-coupled pair to ensure successful writing.

The main issue with the original SAFF is it being too slow in subthreshold, taking up to 10% of a clock cycle. By replacing the weak LL-HVT NMOS by stronger LL transistors, this issue is solved. Alternatively, the complementary topology can be used, shown in fig. 7f [16]. It puts emphasis on the PMOS transistors, exploiting the unbalance that exists in subthreshold to its advantage. The sense-amplifier stage uses a PMOS input pair and a PMOS feeding it from the supply when Clk=1.

B. Comparison

To compare the flip-flops' delay and energy per operation, an input signal with $\alpha = 0.25$ is assumed. The delay t_{d-q} is defined as $t_{d-q} = t_{setup} + t_{clk-q}$. In turn, t_{setup} is defined as the setup time which minimizes t_{d-q} , as suggested in [17].

Fig. 7a compares all topologies for V_{dd} ranging from 0.2 V to 0.6 V. The SAFF-PMOS is the superior option, having both lower energy and lower delay than the other topologies at the lower supply range. The Transmission Gate flip-flop shows comparable but slightly worse performance.

The energy is further decomposed in E_{leak} and E_{dyn} in fig. 7b at the supply implied by the gates. The HLFF and SAFF(-PMOS) have higher dynamic consumption by having every cycle the precharge and clock generation in the former, and the reset in the latter. The SAFF-PMOS has the overall lowest energy due to its lower leakage. The emphasis on the PMOS-transistors makes it well-suited for operating in the unbalanced subthreshold-region of the considered technology at low frequency and/or activity.

Functionality and speed are evaluated under the different process corners to ensure correct operation under process variation. This is done at 0.27 V with a load of 16 inverters. The worst-case t_{d-q} of the SAFF-PMOS is 23.7 μ s in the ss-corner, or 2.4% of the clock cycle. The average energy consumption under typical conditions is 1.43 fJ per clock cycle, of which 0.96 fJ is leakage. Figure 7 also gives the area, which is estimated by summing the $W \cdot L$ -products of all transistors.

VI. COMPARISON WITH A COMMERCIAL LIBRARY

A. Methodology and Benchmarking System

The designed gates and flip-flop form a minimal but complete SCL. This section compares the cells to a commercial library. An always-on feature-extraction system calculating running average and variance has been synthesized using the commercial library. This system is simple but representative for the signal processing tasks which an always-on system could carry out. Table III lists the cells composing this system. The synthesis has been constraint to only use cells of which an equivalent cell is designed.

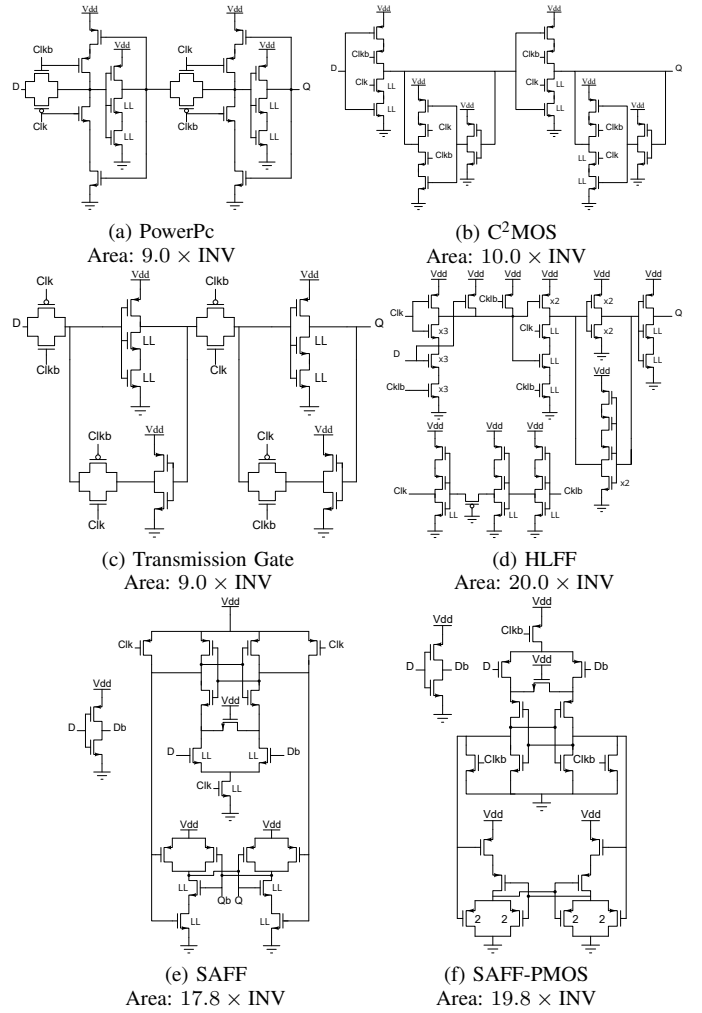


Fig. 7. Flip-Flop topologies modified for subthreshold operation. The area relative to an inverter balanced for superthreshold operation is also given.

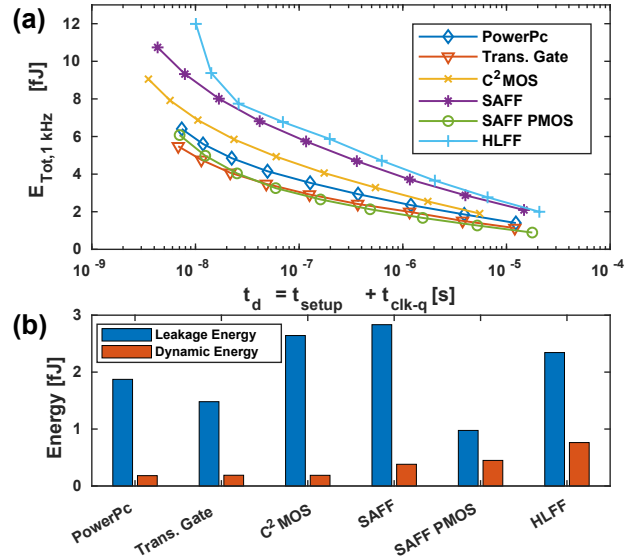


Fig. 8. (a) Comparison of modified flip-flop topologies in energy per operation and delay for V_{dd} varying from 0.2 V to 0.6 V. (b) comparison of E_{leak} and E_{dyn} for the selected flip-flop topologies at 0.27 V.

TABLE III
CELLS COMPOSING THE ALWAYS-ON SYSTEM FOR COMPARISON.

Name	Inverter	NAND2	NOR2	D flip-flop
Amount	930	2227	361	46

TABLE IV
COMPARISON OF THE POWER CONSUMPTION OF THE DESIGNED CELLS
VERSUS COMMERCIAL CELLS FOR THE BENCHMARKING SYSTEM AT 1 kHz.

Library	Leakage Power [pW]	Dynamic Power [pW]	Total Power [pW]
Commercial cells at 1 V	17 065	334.94	17 400
Commercial cells scaled to 0.27 V	1843	24.42	1867
Designed cells at 0.27 V	83.21	78.98	162.2

From circuit simulation results, the power consumption of the proposed cells is estimated by summing the average consumption of each equivalent cell. This omits the capacitance due to local and global interconnect, and thus underestimates the dynamic power. While interconnect is a significant part of dynamic power in recent technology nodes, it stays in the same order of magnitude [18]. Regarding delay, the wire RC-delay is negligible compared to the subthreshold gate delay.

B. Results

Table IV summarizes the power consumption of the system with the newly developed and with the commercial cells. The commercial cells are considered twice, once at 1 V, their nominal voltage, and once at 0.27 V using modeled scale values. This provides an insight in the leakage power reduction achieved by the techniques apart from voltage scaling. The commercial cells result clearly in a leakage power dominated design, with over 98% leakage. The designed cells on the other hand achieve a good balance between dynamic and leakage energy, a sign of operating near MEP. A reduction in total power by a factor 11 is obtained compared to the commercial cells at 0.27 V. Compared to 1 V, the reduction amounts to a factor of 107.

VII. CONCLUSION

This work has proposed a standard cell library (SCL) which optimizes energy consumption in always-on systems targeting 1 kHz operation. Deployed in a representative sensory processing system operating at 0.27 V, it achieves a reduction of two orders of magnitudes in power consumption when compared to a traditional SCL at nominal voltage, or one order of magnitude to the traditional SCL at 0.27 V. A design methodology is developed to achieve these results, which takes into account the constrained operating frequency of always-on systems in contrast to classical computing systems. Both an ideal V_{dd} and V_{Th} are required to achieve minimum

energy, yet V_{Th} of typical CMOS processes is too low for the extremely low target frequency. Leakage reducing circuit techniques are therefore employed to mimic the effect of a higher V_{Th} . The proposed gates succeed in balancing dynamic and leakage energy by combining stacking, body biasing and optimal transistor dimensions relying on second order effect on V_{Th} . Furthermore, flip-flops are examined and modified to operate reliably along with the gates at 0.27 V. The SAFF-PMOS proves to be the best topology, exploiting the unbalance between NMOS and PMOS in subthreshold to its advantage.

REFERENCES

- [1] S. Jeong *et al.*, "A 12nm always-on acoustic sensing and object recognition microsystem using frequency-domain feature extraction and SVM classification," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2017, pp. 362–363.
- [2] B. Warneke, M. Last, B. Liebowitz, and K. S. J. Pister, "Smart Dust: communicating with a cubic-millimeter computer," *Computer*, vol. 34, no. 1, pp. 44–51, Jan 2001.
- [3] R. Sarpeshkar *et al.*, "An ultra-low-power programmable analog bionic ear processor," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 4, pp. 711–727, April 2005.
- [4] B. H. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1778–1786, Sept 2005.
- [5] T. Kunikiyo, K. Mitsui, M. Fujinaga, T. Uchida, and N. Kotani, "Reverse short-channel effect due to lateral diffusion of point-defect induced by source/drain ion implantation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, no. 4, pp. 507–514, Apr 1994.
- [6] S. Sze and M. Lee, *Semiconductor devices: physics and technology*, 3rd ed. Singapore: Wiley, 2013.
- [7] L. A. Akers, "The inverse-narrow-width effect," *IEEE Electron Device Letters*, vol. 7, no. 7, pp. 419–421, Jul 1986.
- [8] M. Alioto, "Ultra-low power VLSI circuit design demystified and explained: A tutorial," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 1, pp. 3–29, Jan 2012.
- [9] S. Narendra, S. Borkar, V. De, D. Antoniadis, and A. Chandrakasan, "Scaling of stack effect and its application for leakage reduction," in *Low Power Electronics and Design, International Symposium on, 2001.*, 2001, pp. 195–200.
- [10] S. G. Narendra and A. Chandrakasan, *Leakage in Nanometer CMOS Technologies*, ser. Series on Integrated Circuits and Systems. Boston, MA: Springer US, 2006.
- [11] N. Reynders and W. Dehaene, *Ultra-Low-Voltage Design of Energy-Efficient Digital Circuits*. Cham: Springer, April 2015.
- [12] G.-Y. W. D. Harris, R. Ho and M. Horowitz, "The fanout-of-4 inverter delay metric." [Online]. Available: <http://pages.hmc.edu/harris/research/FO4.pdf> [Accessed: 16-May-2018].
- [13] G. Gerosa *et al.*, "A 2.2 W, 80 MHz superscalar RISC microprocessor," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 12, pp. 1440–1454, Dec 1994.
- [14] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital integrated circuits: a design perspective*, second edition ed., ser. Prentice Hall electronics and VLSI series. Upper Saddle River: Prentice Hall, 2003.
- [15] H. Partovi *et al.*, "Flow-through latch and edge-triggered flip-flop hybrid elements," in *1996 IEEE International Solid-State Circuits Conference. Digest of Technical Papers, ISSCC, Feb 1996*, pp. 138–139.
- [16] M. Matsui *et al.*, "A 200 MHz 13 mm² 2-D DCT macrocell using sense-amplifying pipeline flip-flop scheme," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 12, pp. 1482–1490, Dec 1994.
- [17] V. Stojanovic and V. G. Oklobdzija, "Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 4, pp. 536–548, April 1999.
- [18] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *Proceedings of the 2004 International Workshop on System Level Interconnect Prediction*, ser. SLIP '04. New York, NY, USA: ACM, 2004, pp. 7–13.