

Subthreshold SCL for Ultra-Low-Power SRAM and Low-Activity-Rate Digital Systems

Armin Tajalli and Yusuf Leblebici
Microelectronic Systems Lab. (LSM)

Ecole Polytechnique Fédéral de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Email: {armin.tajalli, yusuf.leblebici}@epfl.ch

Abstract—The power efficiency of source-coupled logic (SCL) topology for implementing ultra-low-power and low-activity-rate circuits is investigated. It is shown that in low-activity-rate circuits, where the subthreshold leakage consumption of conventional CMOS circuits is more pronounced, subthreshold SCL (STSCL) can be used effectively for reducing the power consumption. An STSCL-based static random-access memory (SRAM) array has been implemented to demonstrate the performance of this topology for ultra-low-power consumption and low-activity-rate digital circuits. A novel 9T memory cell has been developed to reduce the stand-by (leakage) current to 10pA/cell while the SRAM array is operating at 2.1MHz clock frequency. The power consumption benefits of the proposed circuit style can be maintained in nanometer CMOS technology nodes.

I. INTRODUCTION

Source-coupled logic (SCL) circuits have been widely used in very high speed applications [1], [2]. Recently, subthreshold SCL (STSCL) circuits for ultra-low-power applications have been introduced [3]. Precise control on the tail bias current of each SCL gate provides this opportunity to reduce the current consumption of each gate to few pico-Amperes [4]. This property is especially interesting for ultra-low-power circuits where the power consumption of conventional static CMOS circuits is limited by the subthreshold leakage current [5]. Low-activity-rate digital systems provide an example (although not the only one) where leakage power consumption is constituting the dominant part of the total system power dissipation. In this type of applications, as will be shown later, STSCL topology can be employed to reduce the circuit power consumption.

To study the performance of STSCL topology and demonstrate the power efficiency of digital systems constructed based on this topology for low-activity-rate applications, this article presents a very low leakage static random access memory (SRAM) structure. The proposed SRAM cell is capable of robust operation with a bias current of 10pA at a supply voltage of $V_{DD}=500\text{mV}$ and measured static noise margin of $>50\text{mV}$, while operating at 2.1MHz.

II. POWER EFFICIENCY OF STSCL TOPOLOGY

It is shown that SCL gates operating with small logic depth and high activity rate exhibit comparable or better power-delay product (PDP) with respect to the CMOS gates, mainly due to their lower output voltage swing [2], [6]. In this article, we

analyze the performance of STSCL topology for the cases that the activity rate of circuit is very low.

The total power consumption of a system constructed of N STSCL gates with supply voltage of V_{DD} is

$$P_{diss,STSCL} = V_{DD} \sum_{i=1}^N I_{SS(i)} \quad (1)$$

where $I_{SS(i)}$ is representing the bias current of i 'th gate. Based on this, the power dissipation of a STSCL-based circuit is constant and independent of the activity rate. Hence, this type of circuits are more power efficient when the circuit activity rate is maximized [4]. The bias current of each individual cell can be determined separately to optimize the power-delay tradeoff as: $I_{SS(i)} = \ln 2 \cdot V_{SW} C_{L(i)} / t_{d(i)}$, where V_{SW} is the voltage swing at the output of STSCL circuits, $C_{L(i)}$ is the capacitive load at the output of the gate, and $t_{d(i)}$ indicates the delay budget for the proposed gate [2]. Here, the delay of each gate is: $t_{d(i)} \approx \ln 2 \times \tau_i = \ln 2 \cdot R_{L(i)} C_{L(i)}$, where, $R_{L(i)}$ is the load resistance of the proposed gate [4].

For reduced activity rates, the power-delay product (PDP) or energy-delay product (EDP) advantage of SCL diminishes, since the static current consumption of the tail source tends to dominate the overall energy balance [6]. Finally, the relationship between the power consumption and operating frequency (f_{op}) in a STSCL-based digital system is

$$P_{diss,STSCL} \approx \ln 2 V_{DD} V_{SW} f_{op} \cdot \sum_{i=1}^N C_{L(i)} N_{L(i)} \quad (2)$$

where $t_{d(i)} \approx 1/(N_{L(i)} f_{op})$ in which $N_{L(i)}$ stands for the logic depth of the block that the proposed gate is in it. The lower limit for STSCL-based circuit power consumption is the stand-by current of the STSCL gates that can be as low as a few pico-Amperes [4].

On the other hand, conventional CMOS topology shows a very good power efficiency for a very wide range of applications and activity rates. This is mainly due to its negligible static power consumption, as long as leakage is not dominant. For nanometer-scale CMOS technologies where the off-state (subthreshold) leakage of each transistor can reach nA-levels, however, the STSCL topology with its controllable tail bias current can offer reduced power consumption well below the subthreshold leakage of CMOS, while maintaining a significant speed advantage over CMOS topologies.

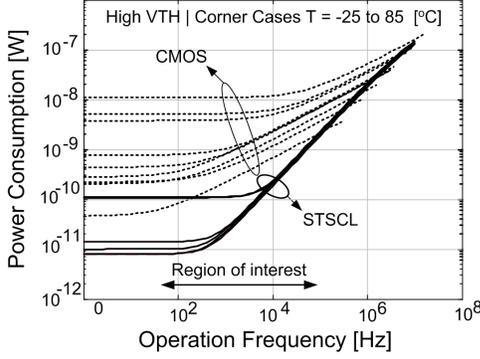


Fig. 1. Simulated power consumption of a chain of gates in 65nm CMOS technology based on static CMOS gates using high V_{TH} transistors (dashed line, $V_{DD,CMOS}=0.3V$) and STSCL topologies (solid line). Variation of the power consumption due to the process corners and temperature variation is shown.

Including leakage current, the total power consumption of a digital CMOS system can be approximated by [7]

$$P_{diss,CMOS} \approx V_{DD} \sqrt{I_{leak}^2 + \gamma \cdot \alpha}. \quad (3)$$

Here, I_{leak} is the total leakage current consumption of the system, α is the activity rate of the system, and γ is a proportionality factor representing the relationship between activity rate and dynamic current consumption of the system. Based on (3), as the activity rate grows, the power dissipation increases proportional to $\sqrt{\alpha}$ in a constant V_{DD} . However, by reducing the activity rate, the power consumption will be dominated by the leakage current: $P_{diss,CMOS}|_{\alpha \rightarrow 0} \approx V_{DD} I_{leak}$.

Comparing (2) and (3) gives the frequency range (or activity rate) in which STSCL topology exhibits better power efficiency. Figure 1 shows the power dissipation of a chain of identical gates based on static CMOS and STSCL topologies in 65nm CMOS technology, both loaded with the same output capacitance. It can be seen that the overall dissipation of the CMOS chain at very low operating frequencies is limited by the leakage current which can be reduced by lowering the supply voltage, yet a dramatic reduction is not possible because the operational robustness diminishes as the current-drive capability of CMOS gates drops exponentially with the supply voltage [8], [9]. Meanwhile, the STSCL topology with a constant tail bias current exhibits comparable operation speed at lower power dissipation, and much less dependence to process and supply voltage variations [7]. Note that the bias current of the STSCL topology can be accurately controlled using high- V_{TH} devices without influencing the speed in weak inversion regime.

III. SRAM CELL DEMONSTRATOR CIRCUIT

Memory circuits can be used to demonstrate the power efficiency of STSCL in low-activity-rate systems. For this purpose, we are presenting an ultra-low-power SRAM array which exhibits very low stand-by dissipation in idle state, and

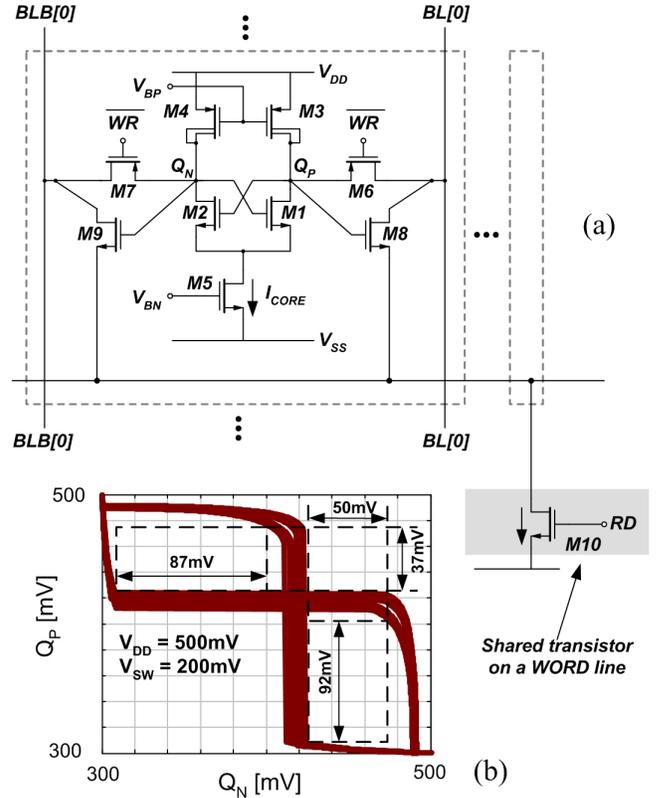


Fig. 2. (a) Completed memory cell. (b) Simulated butterfly curve of a cell in CMOS 65nm (showing different corner cases).

allows robust read and write operations at frequencies that are significantly higher than those achievable in CMOS-based topologies.

The core of the proposed memory cell is based on a cross-coupled STSCL inverter to construct the positive feedback needed to store the data. Circuit schematic of the core of the proposed memory cell is shown in Fig. 2(a). Here, M1 and M2 construct the NMOS switching network, M3 and M4 are the load devices, and the tail bias current is controlled by M5 [3]. A replica bias circuit generates proper bias voltage for PMOS and NMOS devices (V_{BN} and V_{BP}) to control the bias current and output voltage swing. Replica bias circuit also compensates for the effects of process variations. To construct the load resistances, M3 and M4 transistors with their bulk shorted to their drain terminals have been used. Using minimum size devices, this structure shows a very high resistivity in a wide voltage swing [3]. Transistors M6 and M7 in this figure are the access transistors.

The write operation is performed by pre-charging BL and BLB nodes to the desired voltage levels, and then turning on the access transistors M6-M7 in order to charge/discharge the output nodes QP and QN of the memory core. After turning off the access transistors, the positive feedback in the cell will preserve the new state. Since QP and QN have been already charged to the intended values, no extra settling

time is required to accomplish the write operation of the cell. Therefore, the write operation is very fast.

To enable a fast read operation, as illustrated in Fig. 2(a), an open-drain differential pair is formed by M8-M9, driven by the tail bias transistor M10 which is external to the cell and shared by the cells on a word-line. During the read cycle, M10 is turned on and conducts the current I_{READ} , which is steered to one of the output branches of BL/BLB depending on the stored data on the core. This output current is detected by a current-mode sense amplifier (SA) and converted to voltage. Therefore, the speed of the read operation is completely independent of the core tail bias current (I_{CORE}) and depends only on I_{READ} as well as the parasitic capacitances at the nodes BL/BLB.

Isolating the speed of RD/WR operation from the "hold" power consumption in the proposed 9T memory cell permits the reduction of the core bias current down to leakage-current levels. The main limitation for further reducing the tail bias current below 10pA is the turn-on current of the forward-biased source-bulk diode of the PMOS load devices. The forward voltage across this diode is equal to the voltage swing at the output of the core, which can be as low as $V_{SW} = 4nU_T \approx 140\text{mV}$ in room temperature (U_T is the thermal voltage) [4]. In this work, the tail bias current has been chosen to be twice of the junction leakage current.

In contrast to conventional CMOS SRAM cells where the speed of operation depends on threshold voltages, high- V_{TH} devices can be used throughout this cell to limit leakage without impacting speed. Since the tail bias current is very low, the NMOS differential pair devices are deeply in weak inversion, and hence: $V_{GS} \approx V_{T0} + nU_T \ln(I_{CORE}/I_0)$ where V_{T0} is the threshold voltage of the device, and $I_0 = 2n(W/L_{eff})U_T^2$. To have a complete current switching in differential pair transistors, it is necessary that gate-source voltage of the turned on transistor remains larger than V_{SW} or $V_{GS} > V_{SW}$. Therefore, using a device with higher threshold voltage can help to satisfy this constraint. Assuming $V_{GS} \geq V_{SW}$, the minimum theoretical achievable supply voltage is: $V_{DD,min} \geq V_{SW} + V_{CS}$, where V_{CS} is the headroom required to keep the tail bias transistor (M5) in saturation region. For very low bias currents, M5 is in subthreshold region, hence $V_{CS} > 4U_T$. Therefore, the minimum supply voltage is about $10U_T$. Measurements show that the circuit supply voltage (including replica bias circuit and the amplifier used in replica bias) can be reduced to 350mV for very low bias currents [4].

Figure 2(b) illustrates the Butterfly curves of the proposed memory cell in different process corners and temperatures. Here, the voltage swing is chosen to be 200mV at the output of the SCL memory cell and supply voltage is 500mV. Simulations show that the supply voltage can be reduced to 350mV without degrading the static noise margin (SNM) of the cell.

Having a good matching between memory cells and replica bias circuit is necessary to guarantee a high enough SNM value. For this purpose, the size of tail bias transistor (M5) and the PMOS load devices (M4 and M5) need to be large

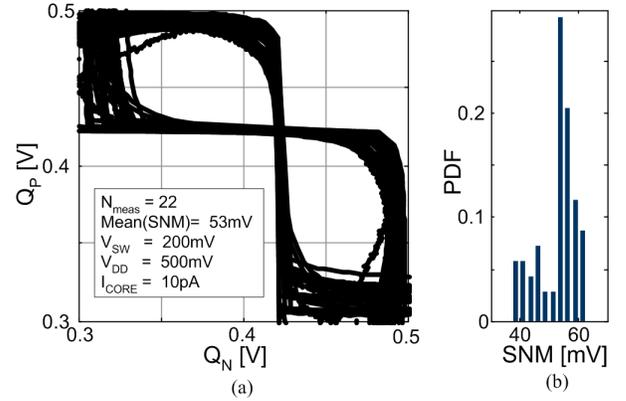


Fig. 3. Measured (a) butterfly curves, and (b) statistical distribution of the SNM (single ended), for the proposed SRAM cell.

enough. As SNM improves by increasing V_{SW} at the output of memory cell, having a high V_{SW} can be useful for improving SNM and hence reducing the size of devices in memory cell.

IV. EXPERIMENTAL RESULTS

A 1kb SRAM array has been designed and fabricated using $0.18\mu\text{m}$ CMOS technology, as a test vehicle to demonstrate the key principles discussed above.

Figure 3(a) shows the measured butterfly curves for the proposed SRAM circuit. The average single ended SNM of the memory cell [Fig. 3(b)] is measured as 53mV for $I_{CORE}=10\text{pA}$, and $V_{SW}=200\text{mV}$. To investigate the influence of V_{SW} on SNM, measurements have been repeated for different output voltage swing values. Figure 4 (top) shows that the SNM initially improves with increasing V_{SW} , and eventually saturates at $V_{SW}=250\text{mV}$, mainly due to the saturation of the amplifier used in replica bias circuits. The dependence of SNM on the tail bias current is shown in Fig. 4 (bottom), with average, minimum and maximum values for SNM plotted for different I_{CORE} levels. It can be seen that the SNM has only minor dependence on I_{CORE} , it remains very stable down to very low levels of bias current, and that the variation on SNM is reduced by increasing I_{CORE} .

In the proposed memory, the main speed limiting factor is the read operation. To increase the speed of operation, it is necessary to increase I_{READ} , which can be achieved by increasing the voltage swing at the gate of M10 in Fig. 2(a). Figure 5(a) shows the variation of the normalized power dissipation of the memory versus operating frequency. With a static current consumption of 10pA/cell, this SRAM core exhibits about three times smaller idle power dissipation compared to [10] while the RD/WR speed can be as high as 2.1MHz (compared to 25kHz for $V_{DD}=350\text{mV}$ in 65nm CMOS technology [10]).

The fabricated 1kb SRAM array is shown in Fig. 5(b). The active area of the memory (including biasing and sense amplifiers) is $670\mu\text{m} \times 390\mu\text{m}$. Measurements confirm that the total current consumption of the array is between 9.5 to

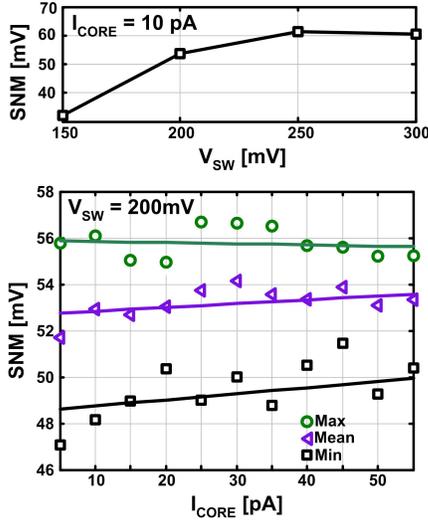


Fig. 4. Measured variation of the SNM versus V_{SW} (for $I_{CORE} = 10\text{pA}$), and variations of SNM versus tail bias current (I_{CORE}) for $V_{SW} = 200\text{mV}$.

13nA for different dies (corresponding to 9 to 12.5pA per SRAM cell) at $V_{DD(SCL)}=500\text{mV}$. At 10pA core bias current and 1.5MHz read/write clock frequency, fewer than 0.01% RD/WR errors were observed. The maximum clock frequency was found between 1.7 to 2.1 MHz for different dies.

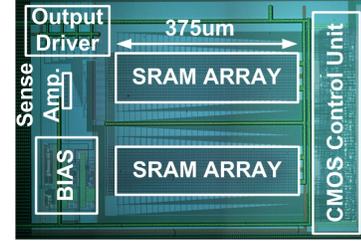
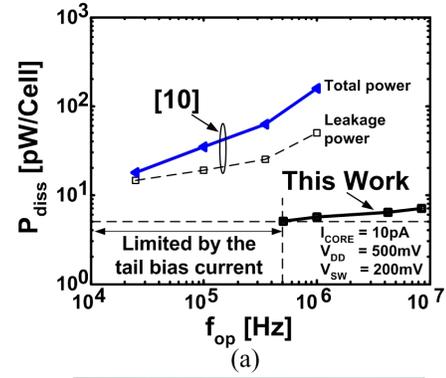
V. OBSERVATIONS AND DISCUSSION

Area and power efficiency of digital CMOS circuits have made them very successful compared to many other types of topologies. The tight tradeoff between the power consumption, speed of operation, supply voltage, and device threshold voltage, however, has made the design of power efficient digital systems in modern nano-scale CMOS technologies very challenging. Some very interesting observations can be made based on the results of this work.

Observation 1: The measurements in this work and also the results in [4] show that the power consumption of each STSCL cell can be reduced to few pico-Watts. Compared to the subthreshold leakage current of CMOS circuits that can be as high as few nano-Amperes per cell, such a low leakage value can be critically important.

Observation 2: It is important to notice that in STSCL circuits, the speed of operation depends on tail bias current of the cells and is independent of the threshold voltage of the MOS devices and also supply voltage as discussed in Section II. In addition, as shown in Section III the minimum supply voltage when the devices are operating deeply in weak inversion does not depend on threshold voltage of MOS devices. Therefore, the tight tradeoff that existed in CMOS topology among supply voltage, threshold voltage, power consumption, and speed of operation, is more relaxed in STSCL.

Observation 3: STSCL topology can exhibit comparable or even better power-delay performance compared to CMOS topology even in low activity rate circuits. This is contrary to the traditional observations that SCL circuits only have been



(b)

Fig. 5. (a) Variation of the idle power consumption (per cell) versus operating frequency, comparing this work with the SRAM cell presented in [10]. (b) The chip photomicrograph of the ultra low stand-by (leakage) current SRAM array (1kb block) fabricated with conventional $0.18\mu\text{m}$ CMOS technology.

used to implement high activity systems [6]. The main reason is that the static power consumption of the CMOS circuits implemented in modern nano scale technologies can no longer be ignored in very low power circuits.

REFERENCES

- [1] P. Heydari and R. Mohanavelu, "Design of ultrahigh-speed low-voltage CMOS CML buffers and latches," *IEEE Tran. on Very Large Scale Integration (VLSI) Syst.*, vol. 12, no. 10, pp.1081-1093, Oct. 2004.
- [2] S. Badel "MOS current-mode logic standard cells for high-speed low-noise applications," *PhD Dissertation*, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, 2008.
- [3] A. Tajalli, E. Vittoz, Y. Leblebici, and E. J. Brauer, "Ultra low power subthreshold MOS current mode logic circuits using a novel load device concept," in *Proc. of Eur. Solid-State Cir. Conf. (ESSCIRC)*, Munich, Germany, Sep. 2007, pp. 281-284.
- [4] A. Tajalli, E. J. Brauer, E. Vittoz, and Y. Leblebici, "Subthreshold source-coupled logic circuits for ultra-low-power applications," *IEEE J. of Solid-State Circuits*, vol. 43, pp. 1699-1710, Jul. 2008.
- [5] M. Pedram and J. Rabaey, *Power Aware Design Methodologies*, Kluwer Academic Publishers, 2002.
- [6] J. M. Musicer, and J. Rabaey, "MOS current mode logic for low power, low noise CORDIC computation in mixed-signal environment," *Proc. of Int. Symp. on Low Power Elect. and Design (ISLPED)*, pp. 102-107, 2000.
- [7] A. Tajalli and Y. Leblebici, "Leakage current reduction using subthreshold source-coupled logic," *IEEE Trans. on Circ. and Syst.-II*, vol. 56, no. 5, pp. 347-351, May 2009.
- [8] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "A variation-tolerant sub-200 mV 6-T subthreshold SRAM," *IEEE J. of Solid-State Circuits*, vol. 43, no. 10, pp. 2338-2348, Oct. 2008.
- [9] B. H. Calhoun, and A. Chandrakasan, "Ultra-dynamic voltage scaling (UDVS) using sub-threshold operation and local voltage dithering," *IEEE J. of Solid-State Circuits*, vol. 41, pp. 238-245, Jan. 2006.
- [10] N. Verma and A. P. Chandrakasan, "A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy" *IEEE J. of Solid-State Circuits*, vol. 43, no. 1, pp. 141-149, Jan. 2008.