Energy-Efficient SRAM Cells for Near-Threshold Computing

354.072 Seminar Mixed-Signal ICs, Summer Term 2021

Severin Jäger

M.Nr. 01613004

severin.jaeger@tuwien.ac.at

Abstract—**Operation at voltages near the transistor threshold allows for computation at minimal energies. However, predominant memory technologies operate poorly in those conditions. To mitigate this bottleneck, several alternative cell designs and technological improvements were proposed in the literature. This seminar thesis discusses selected publication from this field to investigate the design space for near-threshold SRAM cells as there is not dominant solution yet.**

I. Introduction

Modern embedded systems require energy-efficient processors. One promising approach is near-threshold computing, where supply voltages are decreased towards the transistor threshold voltage. This reduces the achievable clock frequencies but increases the energy efficiency far beyond this reduction. However, the static random access memory (SRAM) cells used in the caches of modern embedded processors are a critical component for this technique as established memory cells do not operate reliably at low voltages [1]. In particular, the standard six transistor SRAM cell suffers from process variations, soft errors caused by external radiation and noise in near-threshold operating conditions.

As caches make up a significant share of the die size of systems-on-a-chip (SoCs), the reduction of voltage and thus energy consumption of this circuitry is of high importance. However, alternative approaches both on the technology and the memory cell level are required to achieve reliable near-threshold SRAM. In particular, SRAM cells are increasingly sensitive to radiation-induced bit-flips as voltages are reduced, lack sufficient read stability in larger memory arrays, or do not exploit the advantages coming with lower voltages sufficiently. Once these challenges are tackled, near-threshold computing might enable extremely efficient embedded systems for instance in medical implants or extreme edge computing.

This seminar thesis investigates recent contributions tackling some of the aforementioned limitations of standard SRAM cells in near-threshold setups. The rest of this work is structured as follows. Section II presents basic concepts of near-threshold computing and SRAM cells. In the following, Sections III to V discuss three recent publications investigating problems with SRAM cells in the context of voltage scaling. Finally, major outcomes are summarised in Section VI.

II. Background

A. Near-Threshold Computing

During the last two decades the so-called Dennard scaling that basically stated that increasing transistor counts come at constant power consumption has come to an end. Modern high-performance chips are frequently limited by their power consumption [2]. Thus, the so-called dark silicon era [1] has started. This means that not all functional units of a chip can be active at the same time to keep the power consumption inside the thermal limits.

In general, the power consumption of a digital integrated circuit can be estimated as [3]

$$P_{total} = \frac{CV^2f}{2} + I_{leak}V$$

with the device capacitance C, the supply voltage V the switching frequency f and the leakage current I_{leak}. The first component is called dynamic power as it depends on the clock frequency and the latter static power.

In order to reduce the power consumption, the voltage can be reduced. However, this affects the gate delay as [2]

$$d_{gate} \propto \frac{1}{f} \propto \frac{V}{(V - V_{th})^\alpha}$$

with the transistor threshold voltage V_{th} and the modelling parameter $\alpha > 1$. So if the supply voltage is reduced, the circuit becomes slower. While the dynamic power decreases significantly, the static power plays an increasing role. This is due to the fact that the leakage current increases as the threshold voltage is reduced together with the supply voltage to maximise the operating frequency.

Several techniques like dynamic voltage and frequency scaling (DVFS), clock gating, and power gating have emerged to find a suitable trade-off between performance and energy consumption. One rather radical approach is near-threshold computing, i.e. reducing the supply voltage to the proximity of the transistor threshold voltage. This is desirable as it allows for the optimum energy per operation. The so-called minimum energy point (MEP) is usually slightly above the threshold and depends on process variations and the runtime conditions [4]. At lower voltages, the achievable clock frequency drops
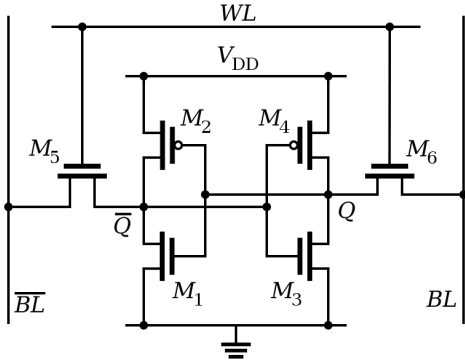
Fig. 1. Standard 6T SRAM cell [9]

Fig. 2. Standard 8T SRAM cell [11]

drastically and increased leakage currents dominate while voltages above the MEP are dominated by dynamic power.

Still, near-threshold computing comes with several design challenges. Firstly, the clock frequency of such a chip is reduced by a factor of 50 to 100 [5]. To utilise this, significant changes in the architecture of the SoC are required. Usually, several parallel execution units are operated simultaneously. This is possible as several parallel cores can operate within the thermal limits of the chip [2]. Furthermore, cell libraries have to be adapted for near-threshold operation [1]. Suitable libraries have however already been developed [6].

The near-threshold computing community aims at maximising the operations per watt of power consumption. Recent designs like the PULP platform [7] target at ultra-low power sensor nodes with notable processing requirements. Such throughput-oriented applications are one promising application field for near-threshold computing, however [5] sees significant potential for this technology in servers and even in personal computing.

B. Static RAM

Typically, static memory is organised in arrays. Each memory cell holds one bit and is connected both to a row address (frequently called word line) and a column address (frequently called bit line). Given a memory address, the row is selected with a decoder. To avoid interference with the non-accessed cells, memory cells have to be in a high-impedance state when they are not selected. In each column, a read-write-amplifier reads or writes the bit lines to or from an output register that allows for fast word-wise read and write operations [8].

The predominant SRAM cell is composed of six transistors and thus called 6T cell. It is depicted in Figure 1 and consists of two cross-coupled inverters (transistors M_1 to M_4) storing one bit of information and the access transistors M_5 and M_6 ensuring the high-impedance state. The cell is constructed differentially, i.e. the voltage at node $\bar{Q}$ is the inverse of the voltage at node Q. In case a logic 1 is stored, the voltage at node Q is almost V_{DD} while the voltage at $\bar{Q}$ is almost 0. There are two differential bit lines, this reduces disturbances [8].
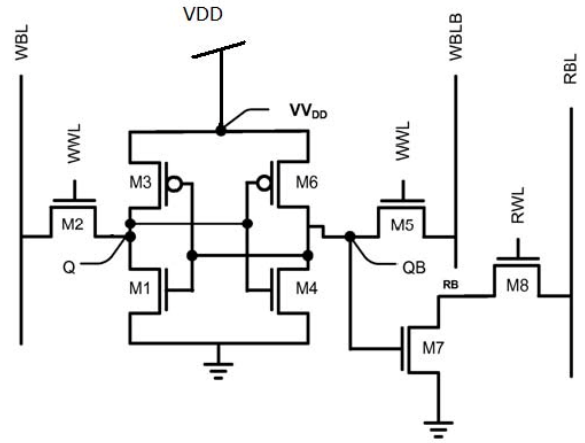
In order to read the information stored in a cell the word line is set high. Both bit-lines are pre-charged to V_{DD}. As one of the nodes Q and $\bar{Q}$ is pulled to V_{DD} and the other one to ground, one of the access transistors (the one on the side pulled to ground) will be conducting, while the other one remains in a non-conducting mode. Assuming the cell stores a 0, M_3 is pulling Q to ground while M_6 is pulling it to V_{DD}. Therefore it is required that M_3 is stronger in the sense that it pulls the whole bit line to ground [10].

The write operation requires the word line to be set high, too. Simultaneously, the write amplifier drives both the bit line and the inverted bit line. Again, the node Q or $\bar{Q}$ is driven both from the cross-coupled inverters and the bit line. This is resolved by the requirement that the access transistors are stronger than the pull-up transistors [10].

So the relative dimensions of the pull-up, pull-down and access transistors have to be well-balanced to allow for suitable read and write performance in standard operating conditions. However, as the noise margins of the 6T cell (read static noise margin (SNM) and write noise margin (WNM)) decrease with the supply voltage [10] and process variations that reduce those margins play an increasing role at near-threshold voltages [5], this cell is hardly applicable for reliable operation at near-threshold voltages.

Dreslinski et al. [5] propose three conceivable ways to overcome this limitations: Alternative SRAM cell designs, robustness analysis in combination with optimisation of the 6T cell and higher-level cache architecture measures. While the latter will not solve the physical limitations of near-threshold SRAM, several transistor and technology level implementations exist in the literature for the earlier approaches.

One well-established alternative cell design is the standard 8T cell proposed in [12]. It is depicted in Figure 2 and decouples the read path from the write path with the additional transistors $M7$ and $M8$. Thus, it introduces read stability, i.e. a read operation cannot flip the stored data [13]. Additionally, in a near-threshold scenario the 8T cell features higher noise margins than the 6T cell and a reduced error rate [14].
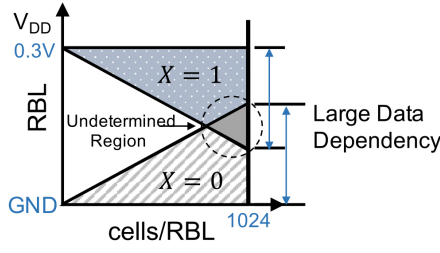
Fig. 3. Reduced effective read bit-line swing in large arrays [15]. Once the undetermined region is reached, the memory will not operate reliably even without noise.

On the other hand the 8T cell comes with additional complexity. Firstly, an additional bit line (RBL) and an additional word line (RWL) are required. Furthermore, the additional transistors obviously increase the cell area and thus decrease the memory density. However, the required optimisations for near-threshold operations increase the size of the 6T cell as well, thus this effect becomes less significant with decreased voltages [5].

III. 10T SRAM CELLS FOR LARGE MEMORY ARRAYS

In [15], three application-specific SRAM cells composed of ten transistors (10T) are proposed. The authors address the limitations of the standard 8T SRAM cell when operated in long arrays (i.e. many words in one block of memory) in a near-threshold configuration.

In particular, the I_{ON}/I_{OFF} ratio of this cell is insufficient for arrays in the magnitude of 1024 cells per read bit line. So the leakage current of the 1023 non-selected cells might be large enough to alter the voltage level on the read bit line in a way that the read amplifier reads the wrong value. Another issue with the 8T cell is the fact that the leakage current depends on the stored data bits. Figure 3 visualises these problems. As a result, topologies consisting of local read bit lines connecting eight SRAM cells which are connected with further circuitry to a global read bit line are used. This additional complexity comes with a significant area overhead.

The authors refer to some alternative cell designs [16–18] but outline their shortcomings regarding area, data-dependent performance and energy consumption. Thus, they propose the three SRAM cells depicted in Figure 4. All three are designed for an improved I_{ON}/I_{OFF} ratio by reducing leakage currents. Thus, they reuse the cross-coupled inverters and the write access transistors from the 8T cell and feature modified read ports. From an array perspective they are except for the 10T-P3 cell which requires an inverted read bit line interchangeable with the 8T cell as they need two words lines (read and write), differential write bit lines, and a single read bit line.

While the 10T-P1 cell is aimed at high performance it still suffers from a data-dependent leakage. In contrast, the 10T-P2 and the 10T-P3 are designed for low power and low area respectively and show a widely data-independent leakage current. In combination with the relative to the 8T
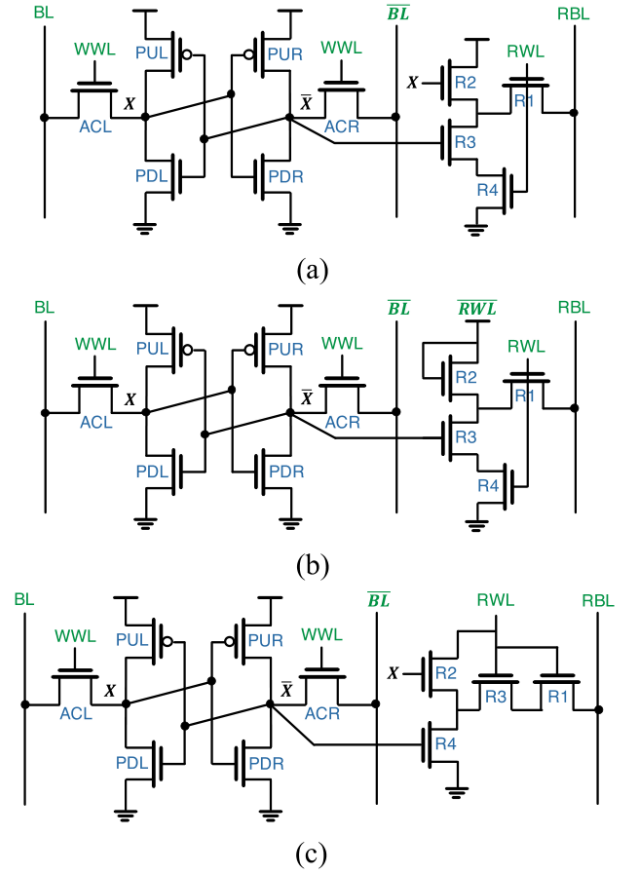

Fig. 4. SRAM cells 10T-P1 (a), 10T-P3 (b), and 10T-P2 (c) proposed in [15]

cell drastically reduced I_{OFF}, they are eligible for long read bit lines.

To evaluate the proposed cell designs, the authors implemented a 128 kB array consisting of four blocks with 1024 words of 32 bits each in a 32 nm process. This was done for all three proposed cells as well as for the reference designs introduced in [16–18]. The following design details were crucial for the performance of the cells: In the 10T-P1 cell, the read transistor $R3$ was designed 1.67 times wider than in all other designs trading off the additional read performance of this cell topology with increased standby power consumption. Furthermore, the inverted read word line had to be provided to the 10T-P3 cell. To minimise the overhead, this signal was shared between two rows. This however leads to half-selected cells that contribute to the overall power consumption.

All designs were simulated and the following results were achieved.

1) Area: All three proposed cells do in contrast to the cells proposed in [16–18] not require any PMOS transistors outside the cross-coupled inverters. This reduces the size of the n-well and leads to narrower cells. This reduces the bit line lengths. Consequently, the dynamic read performance is improved and the dynamic power consumption is reduced. Despite all three proposed cells being smaller than the reference ones, the 10T-P3 cell is clearly the most compact design.

2) Read Bit-Line Swing: The read bit-line swing is quantified as a fraction of the difference between V_{DD} and ground. The simulation shows that the 10T-P1 cell outperforms previous designs, but shows a data dependent bit-line swing. In contrast, the T10-P2 and T10-P3 cells do not show notable data dependencies, however the earlier outperforms all other evaluated cells clearly as it reaches almost $100\,\%$ in most evaluated scenarios. This indicates a minimal leakage current.

3) Standby Leakage Power: Regarding the consumed power while holding the data, all proposed cells outperform the reference cells in the near-threshold region. However this metric itself is insufficient, as SRAM is used in caches and thus frequently accessed.

4) Energy per Access: The authors evaluated the energy consumed by one read access as a function of supply voltage and activity factor (i.e. the fraction of clock cycles in which the SRAM is accessed). They observed that the minimum energy point lies in the near-threshold area for high activity factor and is shifted to slightly higher voltages as the activity is reduced. The 10T-P2 cell outperforms the other cells for all investigated activity factors.

To sum up, the authors have presented three rather specific SRAM cells that are designed with applications in large arrays in mind. One cell has an excellent minimal access voltage, another one features a very low energy per access and a third cell shows a small footprint. All cells outperform previous designs. Still, there is no cell unifying the advantages of all three cells available, thus a careful cell selection depending on the application is unavoidable.

IV. One-Sided Schmitt-Trigger-Based 9T SRAM Cell

In a near-threshold setup, the charge held by one SRAM cell is significantly smaller than in regular super-threshold operation. Thus, the probability of bit-flips induced by alpha particles or high-energy radiation is higher and measures against these so-called soft errors are of great importance. Error-correcting codes (ECC) are a well-established technique to mitigate soft errors in the memory system. However, it requires additional hardware and thus area. A common trade-off is that the ECC hardware is able to correct one bit per word [19].

As the bit-error probability is relatively high in a near-threshold regime and as the memory cells are very small, it is likely that two or more neighbouring SRAM cells are affected by one soft error. Frequently, bit-interleaving is used to avoid multi-bit errors that cannot be resolved by simple ECCs. In a bit-interleaving configuration, words within one row are not stored sequentially but interleaved in the sense that firstly the bits with index 0 of all words are stored then the ones with index 1 and so on [19]. This drastically reduces the probability of multi-bit errors within a word. However, this requires that multiple word are stored in one row. As all of them are connected to the same word select signal and simultaneous read and write operations within one row are desirable, read operations overlapping with write operations in
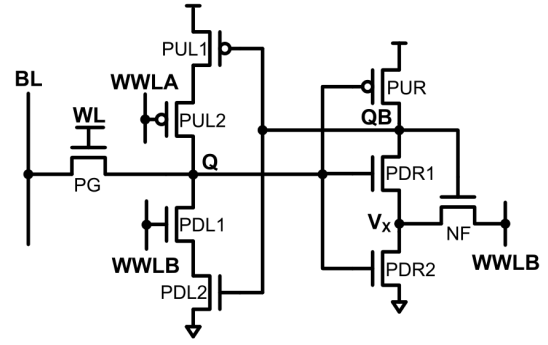

Fig. 5. 9T SRAM cell proposed in [13]

the same row might be disturbed [13]. Common SRAM cells like the 8T cell require some form of write-back mechanism that consumes area and energy to resolve this issue.

Thus, Cho et al. [13] evaluated existing cell designs that do not require such a write-back and found that they either use a differential bit line for read and write and consume significant area and energy or are rather complicated and hence have a large die footprint. Therefore, they propose a novel nine-transistor cell that is depicted in Figure 5. It is based on an inverter ($PUL1$, $PDL2$) with gating transistors ($PUL2$, $PDL1$) cross-coupled with a Schmitt Trigger inverter (PUR, $PDR1$, $PDR2$, NF). Similarly to the 6T cell, he information itself is stored in the nodes Q and QB. Additionally, the access transistor PG connects the memory cell to the single bit line. Additionally, the cell is connected to a row-based word line signal and two column-based word line write signals WWLA and WWLB.

In order to read the value stored in a cell, the word line signal has to be set while the WWLA signal is set to 0 and the WWLB signal is set to 1. This ensures that the left inverter drives the Q node and that this voltage is propagated to the bit line via the PG transistor. One major advantage of the Schmitt Trigger based approach is the higher robustness against disturbances coming from the bit line. Due to the hysteresis of the Schmitt Trigger, the static read noise margin is significantly increased relative to cells based on two standard CMOS inverters or even two Schmitt Trigger inverters.

The Schmitt Trigger based cell comes at the price of increased write complexity. The write process differs between writing 0 and writing 1. In the former case, the signals WL and WWLB are set. After the 0 is driven to the bit line, the WWLA signal is set for a relatively short time to power gate node Q and flip the Schmitt Trigger inverter. In case a 1 is written, WL is set and WWLA remains 0. After the write driver has set the bit line to 1, WWLB is set to 0 for a short time. This disconnects Q from the ground. During this time, the feedback mechanism of the Schmitt Trigger is removed, thus the increased robustness against write processes is not given. Still, the write-1 ability is insufficient. Thus, the WWLB signal is not only driven to 0 but even to a slightly negative voltage. This reduces the trip voltage of the Schmitt Trigger

and simplifies write operations.

One major drawback of this cell is the presence of half-select issues when the cell is operated in a bit-interleaving fashion as it is not read-disturbance free. Due to the good read stability of the cell, there are no major problems with row half-selected cells (connected to the same WL), but this is not the case for column half-selected signals (connected to the same WWLA and WWLB). As the transistors $PUL2$ and $PDL1$ power gate the Q node, the half-selected cells can be floating while the selected cells are written. As WWLA and WWLB are toggled only for a short period, the capacitance of node Q is large enough to maintain the stored value.

To evaluate the design, the proposed cell and some alternative approaches were drawn in a $22\,\text{nm}$ FinFET process and simulated with a Monte Carlo method. The following results were achieved.

1) Area: The proposed cell is $24\,\%$ larger than the standard 8T cell, but still significantly smaller than comparable near-threshold cells without the necessity for write-back.

2) Read Stability: The Schmitt Trigger based 9T cell is not read-disturbance free, yet it outperforms other cells without read-disturbances. Thus, it can be operated at lower voltages without losing sufficient read stability than comparable designs.

3) Energy per Operation: One major advantage of the proposed cell is its low read energy consumption due to its single-ended bit line. Here it outperforms the reference designs. In contrast, it has a relatively high write energy consumption. However, as read accesses dominate in caches, its overall energy per operation in a realistic setting is competitive.

4) Energy-Delay-Product: At the optimal voltage, the proposed cell has a higher energy-delay-product than the other evaluated cells. This makes it extremely efficient for near-threshold computing.

Overall, the one-sided Schmitt Trigger 9T cell is compact and energy-efficient and features a high static read noise margin. On the other hand, it is not read-disturbance free and requires a relatively complex select logic with very specific timing constraints. Still, it is applicable for bit interleaving without write-back and thus for reliable near-threshold SRAM blocks.

V. Near-Threshold Operation of 8T FinFET SRAM Cells

During the last years, FinFET transistors have gained significant prevalence in digital integrated circuits. They resolve the scaling issues regarding standby power and device variability of bulk-CMOS devices by offering better control over the channel and the option to bias front and back gates independently. However, they tend to consume more area than bulk devices at a comparable technology node.

FinFETs are a particularly promising technology for SRAM cells as leakage power plays a notable role in memory power consumption. Furthermore, they allow for reduced delay as the I_{ON} current increases. Overall, their I_{ON}/I_{OFF} ratio is significantly increased as compared to bulk-CMOS cells. As

discussed in Section III, this enables high read static noise margins and long memory columns. Besides the aforementioned area overhead, FinFET SRAM potentially faces data stability issues due to the discrete fin number, however this is not treated further in this work.

Turi and Delgado-Frias [20] are the first ones to evaluate the potential of FinFETs for 8T SRAM cells both in full-V_{DD} and near-threshold conditions. They compare different biasing strategies for the transistors and seek for optimal configurations. They mainly focus on the option of biasing the FinFET back gates. In particular they distinguish two operation modes:

- Shorted gate (SG): In this configuration, back and front gates are shorted and there is only one gate terminal. This corresponds to the typical operation mode of bulk-CMOS transistors.
- Low power (LP): Here, the front and back gates are independent terminals. This is used to reverse-bias the back gate[1] to reduce the leakage current significantly while simultaneously reducing the on-current. However, it requires additional voltage generation circuitry and wiring.

As the paper focusses on the 8T cell, the authors figured out eight different configuration combinations as they mapped different transistors to those operation modes. The read port (transistors $M7$ and $M8$ in Figure 2) is implemented in the SG mode in all combinations to maintain fast read operations while the following transistor pairs are evaluated in all different combinations of SG and LP:

- Access transistors ($M2$ and $M5$),
- N-type inverter transistors ($M1$ and $M4$),
- P-type inverter transistors ($M3$ and $M6$).

For evaluation, two 6T cells (all transistors LP and all transistors SG) and the manifold of 8T variants were implemented in a silicon-on-insulator process with $30\,\text{nm}$ gate length. The memory array for evaluation is as in [15] 1024 words of 32 bits large. All analyses were conducted in a Spice-based simulation environment.

Firstly, the different cell designs were evaluated at full V_{DD}, i.e. $1\,\text{V}$. It was observed that almost all 8T variants outperformed the 6T cells in terms of the energy-delay-product (EDP). In general, the authors concluded that the 8T cell allows for easier configuration for good performance than the 6T cell which has to trade-off read and write performance. Furthermore, the LP configuration with reverse biasing drastically decreased leakage currents both in 6T and 8T designs (by up to $97\,\%$). Interestingly, the best energy-delay-product at this voltage was achieved by the LP_INV cell which uses reverse-biasing (LP) in both the n-type and p-type inverter transistors but SG in the access transistors. It has $74\,\%$ less EDP than the 6T cell in SG configuration.

When it comes to near-threshold operation around the minimum-energy point (at $0.6\,\text{V}$), the 6T LP cell has a higher

---

[1]For n-type transistors a negative voltage (e.g. $-0.2\,\text{V}$) is used, in the case of p-type devices a voltage larger than V_{DD} (e.g. $V_{DD} + 0.2\,\text{V}$) is applied.

energy-delay-product than the SG variant as the read delay of the latter skyrockets already above the threshold voltage. Still, both 6T designs operate with higher energy-efficiency than at full V_{DD}. The 8T cells in contrast show a less steep increase in delay and thus profit from voltage scaling even more. Again, the LP_INV cell achieves the best EDP of $9\,\mathrm{ps}\cdot\mathrm{fJ}$, this corresponds to a $65\,\%$ decrease as compared to full-V_{DD} operation and a $84\,\%$ reduction relative to the best 6T design in the near-threshold regime. The leakage current hardly differs between 6T and 8T cells, however LP designs perform more than one magnitude better than SG cells. In general, the leakage current decreases linearly with the supply voltage. When it comes to noise margins, a decrease is observed with reduced supply voltages. Again, the LP 8T cells show the highest margins in the near-threshold regime. The only performance demerit of the LP designs is the reduced write speed. However, as discussed in Section IV, read accesses dominate in cache applications. Thus, this does hardly limit the potential of this technology.

In the last sections, the authors investigated the performance of the studied FinFET SRAM designs in different memory arrays, unter parameter and voltage variations, and in different temperatures. This however goes beyond the scope of this seminar thesis.

In conclusion, the authors have shown the potential of Fin-FETs for SRAM cells in near-threshold applications. Mainly, they outlined the advantages of both the 8T cell and the low-power configuration with reverse back-gate biasing. Specifically, they pointed out that the LP_INV cell with reverse-biased back-gates in the inverter transistors is a promising candidate for energy-efficient SRAM as it has a minimal leakage current, excellent static noise margins, and an unbeaten energy-delay-product both in the full-V_{DD} and in the near-threshold operation. Thus, FinFETs might offer reliable near-threshold SRAM. However, throughout the publication the area of the proposed cells was not elaborated on. Thus, it remains unclear whether FinFET-based SRAM can compete with bulk-CMOS memory when it comes to integration density.

VI. CONCLUSION

This seminar thesis discussed the challenges for SRAM design in near-threshold computing. As the well-studied 6T cell delivers insufficient performance at these voltages, there is room for alternative solutions. The standard 8T cell is one promising candidate, however it might have a too high leakage current for large arrays or cause write-back overhead in bit-interleaving scenarios. These issues were discussed in the publications discussed in Sections III and IV, which proposed SRAM cells partially mitigating these shortcomings. As none of the discussed cells solves all relevant challenges in near-threshold SRAM design, it is likely that there will be further progress in this field in the near future. Another conceivable development is that technological advances like FinFETs will allow sufficient performance of existing cell designs like the 8T cell as discussed in Section V.

As near-threshold computing is still an emerging field, there are several developments ongoing and the satisfying industry standard for SRAM in this context has not established yet. Thus, cell selection will remain an application-specific trade-off in the near future as several interesting designs including the ones presented in this thesis with their strengths and weaknesses were proposed by academia in the last years.

REFERENCES

[1] V. De, S. Vangal, and R. Krishnamurthy, "Near Threshold Voltage (NTV) Computing: Computing in the Dark Silicon Era," *IEEE Design & Test*, vol. 34, no. 2, pp. 24–30, Apr. 2017.

[2] N. Pinckney, S. Jeloka, R. Dreslinski, T. Mudge, D. Sylvester, D. Blaauw, L. Shifren, B. Cline, and S. Sinha, "Impact of FinFET on Near-Threshold Voltage Scalability," *IEEE Design & Test*, vol. 34, no. 2, pp. 31–38, Apr. 2017.

[3] M. J. Flynn and W. Luk, *Computer System Design*. John Wiley & Sons, Aug. 2011.

[4] M. S. Golanbari, S. Kiamehr, F. Oboril, A. Gebregiorgis, and M. B. Tahoori, "Achieving Energy Efficiency for Near-Threshold Circuits Through Postfabrication Calibration and Adaptation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 2, pp. 443–455, Feb. 2020.

[5] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits," *Proceedings of the IEEE*, vol. 98, no. 2, pp. 253–266, Feb. 2010.

[6] J. Zhou, S. Jayapal, B. Busze, L. Huang, and J. Stuyt, "A 40 nm Dual-Width Standard Cell Library for Near/Sub-Threshold Operation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 11, pp. 2569–2577, Nov. 2012.

[7] D. Rossi, A. Pullini, I. Loi, M. Gautschi, F. K. Gurkaynak, A. Teman, J. Constantin, A. Burg, I. Miro-Panades, E. Beigne, F. Clermidy, P. Flatresse, and L. Benini, "Energy-Efficient Near-Threshold Parallel Computing: The PULPv2 Cluster," *IEEE Micro*, vol. 37, no. 5, pp. 20–31, Sep. 2017.

[8] U. Tietze, C. Schenk, and E. Gamm, *Halbleiter-Schaltungstechnik*, 16th ed. Springer-Verlag GmbH, Jul. 2019.

[9] Inductiveload, *SRAM Cell (6 Transistors)*, https : / / commons . wikimedia . org / wiki / File : SRAM_Cell_(6_ Transistors).svg, Wikimedia Commons, Jan. 2009.

[10] R. E. Senousy, S. Ibrahim, and W. Anis, "Stability analysis and design methodology of near-threshold 6T SRAM cells," in *2016 28th International Conference on Microelectronics (ICM)*, IEEE, Dec. 2016.

[11] D. Tripathy, T. Manasneha, and V. Das, "A single-ended TG based 8t SRAM cell with increased data stability and less delay," in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, IEEE, May 2017.

[12] L. Chang, R. K. Montoye, Y. Nakamura, K. A. Batson, R. J. Eickemeyer, R. H. Dennard, W. Haensch, and D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 956–963, Apr. 2008.

[13] K. Cho, J. Park, T. W. Oh, and S.-O. Jung, "One-Sided Schmitt-Trigger-Based 9T SRAM Cell for Near-Threshold Operation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 5, pp. 1551–1561, May 2020.

[14] A. Gebregiorgis, R. Bishnoi, and M. B. Tahoori, "A Comprehensive Reliability Analysis Framework for

[20] M. A. Turi and J. G. Delgado-Frias, "Full-VDD and near-threshold performance of 8T FinFET SRAM cells," *Integration*, vol. 57, pp. 169–183, Mar. 2017.

NTC Caches: A System to Device Approach," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 38, no. 3, pp. 439–452, Mar. 2019.

[15] S. Gupta, K. Gupta, B. H. Calhoun, and N. Pandey, "Low-Power Near-Threshold 10T SRAM Bit Cells With Enhanced Data-Independent Read Port Leakage for Array Augmentation in 32-nm CMOS," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 3, pp. 978–988, Mar. 2019.

[16] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 3, pp. 680–688, Mar. 2007.

[17] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, "A 0.2 v, 480 kb subthreshold SRAM with 1 k cells per bitline for ultra-low-voltage computing," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 518–529, 2008.

[18] G. Pasandi and S. M. Fakhraie, "A 256-kb 9t near-threshold SRAM with 1k cells per bitline and enhanced write and read operations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 11, pp. 2438–2446, Nov. 2015.

[19] D. Nayak, D. P. Acharya, P. K. Rout, and U. Nanda, "A high stable 8t-SRAM with bit interleaving capability for minimization of soft error rate," *Microelectronics Journal*, vol. 73, pp. 43–51, Mar. 2018.