Energy-Efficient 4T-based SRAM Bitcell for Ultra-Low-Voltage Operations in 28nm 3D CoolCube™ Technology

Reda Boumchedda, Jean-Philippe Noel, Bastien Giraud, Adam Makosiej, Marco Antonio Rios, Eduardo Esmahlotto, Emilien Bourde-Cicé, Mathis Bellet, David Turgis and Edith Beigne
Energy-Efficient 4T SRAM Bitcell with 2T Read-Port for Ultra-Low-Voltage Operations in 28nm 3D Monolithic CoolCube™ Technology

Reda Boumchedda1,2, Jean-Philippe Noel2, Bastien Giraud2, Adam Makosiej2, Marco Antonio Rios2, Eduardo Esmanhotto2, Emilien Bourde-Cicc2, Mathis Belle2, David Turgis1 and Edith Beigne2

1STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles, France
2 Univ. Grenoble Alpes, CEA, LETI, MINATEC Campus, Grenoble, France

ABSTRACT

This paper presents a 4T-based SRAM bitcell optimized both for write and read operations at ultra-low voltage (ULV). The proposed bitcell is designed to respond to the requirements of energy constrained systems, as in the case of most of the IoT-oriented circuits and applications. The use of 3D CoolCube™ technology enables the design of a stable 4T SRAM bitcell by using data-dependent back biasing. The proposed bitcell architecture provides a major reduction of the write operation energy consumption compared to a conventional 6T bitcell. A dedicated read port coupled to a virtual GND (VGND) ensures a full functionality at ULV of read operations. Simulation results show reliable operations down to 0.35 V close to six sigma (6σ) without any assist techniques (e.g. negative bitlines), achieving in worst case corner 300 ns and 125 ns in write and read access time, respectively. A 6x energy consumption reduction compared to a ULV ultra-low-leakage (ULL) 6T bitcell is demonstrated.

1 INTRODUCTION

In IoT applications, the chip works with limited energy source stored in an embedded battery. Therefore, the power management is an important factor and must be optimized to improve the battery life time. An efficient way to reduce both active and standby power is to lower the supply voltage, while keeping good performance in the active mode.

In many recent applications, the embedded SRAM macro may occupy more than 50% of SoC total area. In consequence, the memory is often the main contributor to the total static power consumption. Moreover, at low voltage operation, the SRAM is the main limiting factor in terms of performance. Improving the energy efficiency while maintaining SRAM performance is therefore the key to enable long-battery-life SoCs for IoT applications.

In modern chips, the 6T bitcell is widely used to design SRAM circuits due to its high stability in retention and performance in write and read operations at nominal voltage. In ULV operation, however, this bitcell suffers from lack of reliability for write and read operation. In write operation (particularly at low temperature) the pass-gate transistor weakens and shows difficulty to toggle the bitcell data. In read operation (particularly at high temperature) the \( I_{ON}/I_{OFF} \) ratio decreases, increasing the difficulty to obtain enough voltage difference to perform a reliable read operation. To overcome these issues, a lot of assist techniques have been developed, such as negative bitlines (NBL) [1] or WL underdrive [2] but at the expense of power consumption and area overhead [3].

On the other side, the 4T bitcell, where either the pull-up (load less) or pull-down (driver less) transistors are removed, could become an interesting alternative for ULV SRAM design. Its main advantage is the easy writability owing to the fact that there is no “fight-back” from the bitcell when toggling the data, as there is no standard latch in the cell. The well-known disadvantage of this bitcell is the difficulty to ensure a balance between sufficient stability in data retention and in read operation. Designing with FD-SOI-based process technologies (e.g. planar FD-SOI or 3D CoolCube™), it has been demonstrated that 4T bitcell can be made stable in retention and read mode by tuning several knobs such as back-plane type and biasing, gate type and length and counter doping[4][5][6][7][8].

Based on previously published work [6][7][8], we propose a 4T-based SRAM bitcell further optimized for ULV operation, particularly for energy-efficient write operation. Moreover, we propose to use a dedicated read port coupled to a VGND [9] to improve even more the performance of read operation. Finally, we design this bitcell in 3D CoolCube™ technology based on a novel SRAM bitcell array organization avoiding word interleaving within a row.

The rest of the paper is organized as follows:

Section II presents the 3D CoolCube™ technology and its advantages in SRAM design. Section III introduces the proposed 3D 4T bitcell. Section IV discusses the simulation results obtained on the 3D 4T bitcell. Section V compares the designed 3D 4T bitcell to a 6T bitcell. Section VI draws conclusions and exposes some perspectives.

2 3D COOLCUBE™ TECHNOLOGY

Among the 3D monolithic technologies, 3D sequential LETI CoolCube™ technology [10] offers a fine 3D interconnect pitch compared to existing technologies. This feature the way for efficient 3D-VLSI circuits, aiming to reduce the congestion on the BEOL while providing real 3D routing possibilities [11]. In the 3D-monolithic technology, transistors are made on several stacked tiers, with a tier being a thinned wafer. On the upper tiers there is the ability to process asymmetric double-gate transistors, resulting in a better on-to-off current ratio. Furthermore, this feature brings new possibilities on design level that allow an innovative design to enhance the stability, power consumption and/or performances.
Figure 1 shows how the bitcell is going to be integrated in 3D. On three superposed tiers, two bitcells are integrated with their read ports put in-between the two 4T bitcells. This approach exhibits a gain in density while saving the addition of an extra tier by sharing one tier for the read ports. In addition, the memory can be designed in a specific way such that there is no word interleaving and thus avoid half-selected [12] bitcells where the retention of the 4T bitcell is unstable.

### 3 PROPOSED 3D 4T + 2T SRAM BITCELL

The proposed bitcell is shown in Fig. 2. The bitcell contains a core of four transistors, to hold and write the data, and a read port of two transistors to read out the data.

The read port used with the 4T is specifically designed to enhance read operation at low voltage where the \( \frac{I_{\text{ON}}}{I_{\text{OFF}}} \) ratio is critical.

To improve readability at low voltage, a virtual ground (VGND) is added to the read port as shown in Fig. 2. Figure 3 shows the case where a read operation is done w/o and w/ VGND. On Fig. 3(a), no VGND is added to the read port. Supposing that all the bitcell in the column contain a ‘0’, the read current is joined by \( N-1 \) leakage currents (\( N \) being the number of rows). At ULV and with a high number of rows, the total leakage current can be equivalent to the read current, resulting in the discharge of the read bitline (RBL) no matter the content of the read bitcell. On Fig 3 (b) with VGND added to the read port the read current is joined by only one leakage current, assuring a distinctive read operation at ULV, high temperature and high number of rows.

Since the reliability of read operation is assured by adding a read port with row based VGND, the 4T core of the bitcell can be optimized only for best balance between retention and write stability. The mechanism of retention in the 4T bitcell relies on the equilibrium of the leakage current present in the bitcell, as detailed in previous works [6] [7].

In 3D monolithic design there is the possibility to use the data-dependent dynamic back bias to enhance the bitcell stability in retention. As shown in Fig. 2, the pull-up (PU) back gates are connected in dynamic threshold-voltage MOS (DTMOS) [8] configuration (back gate/plan of one MOS connected to its front gate) whereas pass gate (PG) back-gates are connected to their opposite storage nodes BL/T and BL/F. This configuration optimize each branch of the 4T bitcell independently. On the branch where the internal node holds a ‘1’ the PU is made stronger than the PG to help the bitcell maintain the ‘1’ at VDD. Intrinsically on the other side, the PG is stronger than the PU so that the ‘0’ contained in the internal node remains close to GND.

An important advantage of the 4T bitcell is its easy writability. Owing to the absence of pull-down transistors in the 4T driver less (4TDL) bitcell, there is no fight-back from the bitcell when written, allowing an easier toggle and with low energy cost. At low supply voltage where the write operation is difficult, a standard 6T bitcell requires the application of write assist techniques that are costly both in terms of silicon area and energy. On the other hand, the 4T bitcell is free from those constraints. This difference is clearly shown on Fig. 4 and Fig. 5. On Fig. 4(a) the light-back of the 6T bitcell against the write operation causes the latches of the bitcell to consume current through short circuit between VDD and GND. On Fig. 4(b) the 4T bitcell exhibit no short circuit current thanks to the removal of the pull-down (PD) transistors connected to GND.
Figure 5(a) shows that the 4T bitcell pulls the internal node to VDD faster than the 6T bitcell. This is due to the fight-back of the 6T bitcell against the write operation as mentioned before. Moreover, this issue causes the 6T bitcell to consume more current than the 4T SRAM bitcell during the write operation as shown on Fig. 5(b).

Fig. 5. Simulation of the (a) rise of the internal node to VDD and (b) current of the 4T and 6T bitcell during a write operation.

4 SIMULATION RESULTS

The simulations were done using the importance sampling method, as available in a commercial simulator [13]. Typically, the SRAM stability is targeted at 6 σ, corresponding to approximately 1 fail per billion. Importance sampling methods enable stability evaluation within this low failure range with good accuracy and low execution time.

Figure 6 shows the yield expected for a SRAM for a given bitcell σ. The yield is given as the number of maximum Mbit where at most one failing bitcell is present. The graph also shows, for a given σ, the number of cuts with at most one failing cut for a given size of the SRAM. As expected, the lower the size of the SRAM the higher the number of cuts. For a 64 kbit SRAM, the number of cuts is 15466 and 4601 at 6σ and 5.8 σ, respectively.

The bitcell is designed using the 28 nm FDSOI models whose model card is slightly adjusted to satisfy all the bitcell metrics. Therefore, to enhance the retention stability, the Vt gap between the NMOS and PMOS is enlarged by 120 mV (+60 mV PMOS Vt, −60 mV NMOS Vt). This optimal balance between NMOS and PMOS Vt is achievable using the available process tuning handles as demonstrated with silicon measurements in 14 nm FDSOI [7].

a) Data retention on 4T SRAM bitcell

In this sub-section the retention stability of the 3D 4T bitcell is investigated. As mentioned before, the retention of the 4T relies on an equilibrium between the PG (NMOS) and the PU (PMOS) transistors. This means that the most critical corners for stability evaluation are FS and SF, corresponding to fast NMOS and slow PMOS and slow NMOS and fast PMOS, respectively. In [6][7] is mentioned that the stability rely on the ratio of resistance between the PU and the PG. This ratio defines the level of the internal node maintained at ‘0’, to be close to GND the resistance of the PU must be higher than the one of the PG.

However, at ULV and with the degradation of the IOFF/IOFF ratio, the leakage current of the PG (saturated ILEAK) can become equivalent to the On current of the PU (linear ION) on the node holding a ‘1’ in the bitcell. In this case the ‘1’ contained in the 4T bitcell can be pulled down by the leakage current of the PG, resulting in bitcell data loss in worst-case corner.

TABLE I. 4TDL RETENTION STABILITY ANALYSIS AT FS AND SF CORNER FOR VARIOUS TEMPERATURES AND SUPPLY VOLTAGES

| VDD (V) | Temperature (°C) |
|--------|------------------|
|        | -40  | 25  | 125 |
| FS    | SF   | FS  | SF  | FS  | SF  |
| 0.38  | 7.6  | 7.2 | 7.5 | 6.9 | 7.1 | 6.0 |
| 0.37  | 7.1  | 7.1 | 7.0 | 6.8 | 6.7 | 6.0 |
| 0.36  | 6.7  | 6.9 | 6.6 | 6.7 | 6.3 | 5.9 |
| 0.35  | -40°C | 6.2 | 6.9 | 6.2 | 6.6 | 5.9 | 5.8 |
| 0.34  | 5.8  | 6.7 | 5.7 | 6.6 | 5.4 | 5.7 |

Table I shows the σ number obtained for the retention stability of the 3D 4T bitcell in the FS and SF corners. For both corners the worst-case stability is located at low voltage and high temperature (125°C). At this temperature the IOFF/IOFF ratio worsen. On the SF corner the leakage of the PG become insufficient to maintain the ‘0’ whereas on the FS corner the leakage of the PG become too strong and pull down the ‘1’ to GND. Through design lever, the stability condition can be attained for both corners. A reliable retention for the 3D 4T bitcell is achievable at a V_MIN of 0.35 V, where a σ of 5.9 and 5.8 is attained for the FS and SF corner, respectively. As shown previously on Fig. 6, the yield does not depend only on the σ but also on the size of the SRAM. In the targeted applications of this SRAM (IoT-oriented circuits and applications), the size of the memory is usually lower than 64 kbit [14][15]. Hence with a 5.8 σ and considering a size of 64 kbit a likely yield of 99.99% is obtained, leading to 1 fail at most out of 4601 cuts.
b) Write operation on 4T SRAM bitcell

This sub-section investigates the write operation which is one of the strong points of the 4T bitcell for the fact that there is no fight-back from the bitcell during the write, this condition allows a write operation at ULV with no assist and low energy cost. The stability of write operation is evaluated in function of the minimum wordline pulse duration (WL pulse) that has to be applied to the bitcell in order to write the cell correctly.

Table II summarizes the results of write stability estimation expressed in #σ for different process corners at VDD=0.35 V and T=25°C. The SS corner (where both the PG and PU are slow) is the worst case regarding writability on the 4T bitcell.

| Process | VDD = 0.35 V | WL Pulse (ns) |
|---------|--------------|---------------|
| T = 25°C |              |               |
| FF      | 15.0         | 14.7          |
| FS      | 12.4         | 12.2          |
| TT      | 13.5         | 13.2          |
| SF      | 14.8         | 14.5          |
| SS      | 11.8         | 11.5          |

Table III depicts the results of a similar analysis as in the previous table but for various temperatures in SS corner. Results exhibit that the temperature has a major impact on the writability, the lower the temperature the more difficult it is to write the bitcell.

| Process = SS | VDD = 0.35 V | WL Pulse (ns) |
|-------------|--------------|---------------|
| Temp (°C)   |              |               |
| 125         | 24.9         | 24.5          |
| 85          | 20.4         | 20.1          |
| 25          | 11.8         | 11.5          |
| 0           | 8.3          | 8.0           |
| -40         | 2.9          | 2.6           |

Table IV shows the writability evaluation results for various supply voltages, at worst-case corner process and temperature of SS and -40°C, respectively.

Even though the 4T bitcell does not need any assist for write operations, the addition of negative bitline (NBL) assist is studied to estimate the speed up factor on the write operation.

Table V depicts the obtained results on write operation when using NBL assist for the worst-case PVT. With only a 10% VDD NBL a gain of almost 4x is obtained reaching up to 30x if 30% NBL is considered.

| Process = SS | T = -40°C | WL Pulse (ns) |
|-------------|-----------|---------------|
| VDD (V)     | 100       | 400           | 300           | 200           | 100           |
| 0.39        | 10.8      | 9.5           | 8.5           | 6.7           | 6.7           |
| 0.38        | 9.9       | 9.4           | 8.6           | 7.6           | 5.8           |
| 0.37        | 9.1       | 8.5           | 7.8           | 6.7           | 5.0           |
| 0.36        | 8.2       | 7.6           | 6.9           | 5.8           | 4.1           |
| 0.35        | 7.3       | 6.8           | 6.0           | 5.0           | 2.9           |

c) Read operation on read port with VGND

At low voltage the conventional read operation is unreliable because of the degradation of the I_on/I_off ratio. The lower the ratio, the more the read current is similar in value to leakage current. In these conditions, difficulty arises to distinguish a ‘1’ from a ‘0’ during read operations. This phenomenon grow stronger at higher temperature and large number of bitcell per column (b/c), due to temperature adversely impacting the I_on/I_off ratio and more leakage contributors for larger column size.

To ensure a functional read operation at low voltage, a read port with a VGND is added to the 4T bitcell. When a bitcell is read in a column, its VGND drops to 0V while the VGND of the remaining read ports of the column remains at VDD. As a result, the leakage of unaccessed cells in the column is suppressed, removing their impact on the bitline discharge in read operation. Only the read port leakage of the adjacent row remains presents, which make one read current facing a leakage current instead of one read current facing (N−1) leakage currents, where N is the number of b/c.
in unaccessed rows charges back the read bitline towards VDD, thus slowing down the read operation. Another observation can be made on the effect of VBB, the greater the # of b/c the greater the effect of VBB is, an important speed up on WL pulse can be seen at 256 b/c, but at 32 b/c the effect is minor. On Fig. 8(a) a read operation is performed at worst-case corner and V_{MIN} (SS[0.35 V]|-40°C) with a WL pulse of 125 ns on 256 b/c.

5 4T VS 6T SRAM BITCELL COMPARISON

In this section the proposed 3D 4T bitcell is compared to a 6T bitcell optimized for operation at ULV with ULL. The reference 6T bitcell is simulated using its respective foundry models. The proposed 3D 4T bitcell however, is simulated using standard logic models.

a) Write operation

The write operation of the 6T bitcell is evaluated at its worst-case corner (SF |-40°C) and at the V_{MIN} of the 3D 4T bitcell. As shown in Table VI, writing the 6T at ULV is not feasible without NBL assist. With at least 30% VDD of NBL, a 6 σ stability in the write operation can be achieved with a WL pulse of 200 ns. A level of 30% NBL (~105 mV) at 0.35 V is difficult to implement and require a large capacitance or a second power supply to be able to generate this negative voltage, representing an important addition of area and energy consumption [3].

Table VII Leakage analysis in cross corner at multiple VDD for the 4T and 6T bitcell.

b) Leakage current

The leakage in the 4T is necessary for the stability of data retention. Using logic models for the 4T and specific low leakage models for the 6T, the 4T bitcell cannot overcome the reference cell in terms of leakage, but the gap can be minimized. Table VII shows the ratio of the leakage between the 4T and the 6T. The worst case is at SS[0.35 V]|-40°C with a ratio of 58, at worst case of leakage (FF[0.35 V]|125°C) the ratio drops to 12. Even with unfavorable leakage compared to the 6T, next section shows how the 4T can overcome this.

Table VII Leakage analysis in cross corner at multiple VDD for the 4T and 6T bitcell.

c) Energy and leakage trade-off

Even though the 4T bitcell possesses a higher leakage as compared to the 6T one, the 4T bitcell gains in total in the write operation. Since the 4T bitcell do not fight-back the toggle in a write operation and does not need using NBL assist, the write
operation is done with low energy expense. On the other hand, the 6T resists the toggle during write operation, requiring for write operation a strong NBL (30% VDD).

Figure 9 shows the accumulation of the energy spent during 1000 write operations of 32 bits word and then the energy lost in leakage while idling. The 6T uses much more energy during writing. After 1000 write operations, the 6T has spent 6 more times more energy than the 4T in the worst case. After a while idling, the leakage of the 4T being higher than the one of the 6T, the energy expense of the 4T catches up the one of the 6T. \( T_{\text{RECOVERY}} \) is defined as the maximum lapse of time where a write operation has to be done so that the energy expense of the 4T is lower than the one of the 6T, in the worst case \( T_{\text{RECOVERY}}= 350 \) cycles (considering writing 50% of the word bitcells).

If an asynchronous system is considered for the proposed 4T bitcell, it benefits at maximum from the bitcell advantages. In an asynchronous system, the cycle time adapts itself to the variation of process, supply voltage and temperature (asynchronous cycle time) \[16\], thus avoiding periods where the SRAM is ready to work but is idling instead while waiting for the next operation (synchronous cycle time). Consequently, the system carry operations on the SRAM as soon as available, and hence make sure that the 4T bitcell energy consumption stay below the one of the 6T bitcell.

6 CONCLUSION

This paper demonstrated the efficiency of the 4T-based SRAM bitcell compared to the 6T-based one to drastically reduce the energy consumption of write operations. The simulations results showed a 6x reduction in the worst-case conditions. This is mainly due to the fact that there is no period of short circuit inside the bitcell when toggling the data. This point is crucial for subthreshold operations where this period can be relatively long due to low \( I_{\text{ON}}/I_{\text{OFF}} \) ratio. To use the proposed bitcell in the optimal write and retention conditions, a dedicated read port is mandatory. Moreover, the bitcell has to be used without word interleaving to avoid the half-selected disturb effect, leading to data loss. To overcome this issue, a new SRAM bitcell array organization has been proposed in 3D CoolCube™ technology. This structure enabled to use the bitcell without word interleaving into the array. This process technology enabled to dynamically use the back gates improving the data retention stability. A 125 ns read time has been showed by simulation in the worst-case condition, while achieving 300 ns of write time without assist technique at 0.35 V. Furthermore, it is demonstrated that a SRAM based on the proposed bitcell is more energy efficient than a 6T bitcell if a write operation is performed at least every 350 cycles.

In a further work we expect to reduce the leakage current of the 4T bitcell, the lower the leakage the longer the period of no activity where the energy consumption of the 4T is lower than the 6T SRAM bitcell.

7 REFERENCES

[1] V. Kumar et al., "A Sub-0.5V Reliability Aware-Negative Bitline Write-Assisted 8T DP-SRAM and WL Strapping Novel Architecture to Counter Dual Pattering Issues in 10nm FinFET," VLSID, pp. 269-274, 2017.
[2] E. Karl et al., "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry," ISSCC, pp. 230-232, 2012.
[3] S. Kumar et al., "SRAM write assist techniques for low power applications," ICSC, pp. 425-430, 2016.
[4] J.-P. Noel et al., "Robust multi-VT 4T SRAM cell in 45nm thin BOx fully-depleted SOI technology with ground plane," TCAD, pp. 191-194, 2009.
[5] V. Asthana et al., "Circuit optimization of 4T, 6T, 8T, 10T SRAM bitcells in 28nm UTBB FD-SOI technology using back-gate bias control," ESSCIRC, pp. 415-418, 2013.
[6] M. Brocard et al., "High density SRAM bitcell architecture in 3D sequential CoolCube™ 14nm technology," S3S, pp. 1-3, 2016.
[7] R. Boumechella et al., “High-Density 4T SRAM Bitcell in 14-nm 3-D CoolCube Technology Exploiting Assist Techniques,” TVLSI, Volume 25, Issue 8, pp. 2296-2306, 2017.
[8] F. Andrieu et al., “Design technology co-optimization of 3D-monolithic standard cells and SRAM exploiting dynamic back-bias for ultra-low-voltage operation,” IEDM, pp. 20.3.1-20.3.4, 2017.
[9] N. Verma et al., "65nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy," ISSNCG, Digest of Technical Papers, pp. 328-329, 2007.
[10] P. Batude et al., "32VL5SI with CoolCube process: An alternative path to scaling", Symposium on VLSI Technology, pp. T48-T49, 2015.
[11] F. Clermidy, et al., "Technology scaling: The CoolCube™ paradigm," S3S, pp. 1-4, 2015.
[12] Z. Chuan Lee et al., “NBTI/PBTI-aware wordline voltage control with no boosted supply for stability improvement of half-selected SRAM cells,” ISQOC, pp. 200-203, 2012.
[13] L. Ciampolini et al., “Efficient yield estimation through generalized importance sampling with application to NBL-assisted SRAM bitcells,” ICCAD, pp. 1-8, 2016.
[14] Y. C. Chien et al., "A 0.2 V 32-Kb 10T SRAM With 41 nW Standby Power for IoT Applications," TCAS I, 2018.
[15] H. Fujiiwara et al., "A 64-Kb 0.37V 28nm 10T-SRAM with mixed- Vth read-port and boosted WL scheme for IoT applications," A-SSCC, pp. 185-188, 2016.
[16] J. Chen et al., “An ultra-low power asynchronous quasi-delay-insensitive (QDI) sub-threshold memory with bit-interleaving and completion detection,” NEWCAS, pp. 117-120, 2010.