Ultra Low Power Litho Friendly Local Assist Circuitry For Variability Resilient 8T SRAM

Vibhu Sharma[1,2] , Stefan Cosemans[1,3]
[1]ESAT-MICAS Laboratory, K.U Leuven, Leuven, Belgium

Maryam Ashouei[2], Jos Huisken[2]
[2]Holst Centre/imec, Eindhoven, Netherlands

Francky Catthoor[3,1] & Wim Dehaene[1,3]
[3]imec, Leuven, Belgium

***Abstract*— This paper presents litho friendly circuit techniques for variability resilient low power 8T SRAM. The new local assist circuitry achieves a state-of-the-art low energy and variability resilient WRITE operation and improves the degraded access speed of SRAM cells at low voltages. Differential VSS bias increases the variability resilience. The physical regularity in the layout of local assist circuitry enables litho optimization thereby reducing the area overhead associated with existing local assist techniques. Statistical simulations in 40nm LP CMOS technology reveals 10x reduction in WRITE energy consumption, 10^3x reduction in write failures, 6.5x improvement in read access time and 31% reduction in the area overhead.**

Keywords- SRAM 8T cell, variation, Write Margin, local write receiver, litho optimized.

I. INTRODUCTION

Read-decoupled 8T SRAM cell offers a higher degree of variability resilience compared to 6T SRAM cell at the expense of an increased area overhead [1]. The 8T SRAM cell area overhead will tail off with technology scaling, advanced technology nodes of 32 nm will witness 8T SRAM cell area comparable to the 6T SRAM cell [2]. The requirement for scaled operating voltages and use of LP technologies for low power design necessitates the use of upsized READ stack transistors. Therefore, comparison of 8T SRAM cell at the constant value of read current would result in an increased size of READ stack transistors, thereby dwindling the scaling of 8T SRAM cell [3]. Secondly the write margin (WM) improvement with read-decoupled 8T SRAM cell at scaled voltage levels is not that significant, compared to the read stability. Thirdly the low swing bit-line reduces the READ energy consumption but the WRITE energy consumption is not optimized as the WRITE operation requires full voltage swing on the bit-lines of an accessed SRAM cell. Therefore WRITE energy consumption is more critical than the READ energy consumption and is a vital issue for realizing ultra low energy SRAMs. The key for reducing WRITE energy lies in reducing the voltage swing on the bit-lines. Half swing during the WRITE operation as proposed in [4] reduces the WRITE energy consumption theoretically by 75%. Further reduction of the voltage swing on the highly capacitive bit-lines is

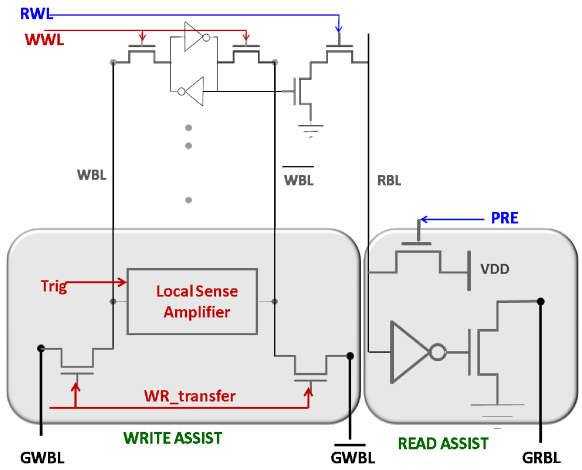


Figure 1. Conventional local sense amplifier and read buffer based local assist circuitry.

limited by the degraded WM for scaled technologies at the lower voltages. SAC SRAM [5] results in 90% reduction in WRITE energy consumption but the area overhead is very high. In addition to the area overhead of SAC-SRAM, the GND connection via NMOS transistor degrades the read SNM and reduces the cell read current. The hierarchical bit-lines with local sense amplifiers [6-8] achieve low energy WRITE operation (Figure 1). The low swing data information on the highly capacitive global bit-lines is first transferred onto the short local bit-lines through NMOS pass transistors. Then the local sense amplifier resolves this low swing information to full voltage level on the low capacitive local bit-lines. The low swing data information on high capacitive global bit-lines and full swing voltage signal on the low capacitive local bit-lines reduces the energy consumption of WRITE operation. The presence of local sense amplifier as local assist in memory requires complex memory matrix optimization and the area overhead is high. It also does not address the write-ability degradation of SRAM cells at the reduced voltage levels.

The size of the transistors in local sense amplifier acting as local write receiver (Figure 1) is dictated by the mismatch offset voltage and the speed requirement. The conventional way of reducing the mismatch offset requires the upsizing of

978-3-9810801-8-6/DATE12/©2012EDAA

the local write receiver transistors [9]. This directly translates into increased energy consumption and the area penalty of the local write receiver. Secondly, the traditional strobed local write receiver requires several critical timing signals that must be applied in a sequential order with sufficient margins. The low swing data is first transferred from the global bit-lines onto the local bit-lines and the local write receiver can only be triggered when the low swing voltage is more than the mismatch offset of the local write receiver. This requires an expensive timing circuitry and also introduces significant timing margins thereby increasing the access time. This paper introduces various circuit techniques to address all above mentioned issues:

1) Novel low-swing write mechanism enables low energy WRITE operation:

a) non strobed local write receiver (NS-LWR) reduces the timing complexity associated with existing state of the art strobed local write receivers (LWR)

b) The differential VSS bias effort on the NS-LWR for offset mitigation compounds into dual action for improving WM of the accessed SRAM cells.

2) Local read buffer compensates degraded Iread and achieves high performance.

3) 8T SRAM cell type structure of the local assist circuitry (NS-LWR, WR MUX and local read buffer) results in litho friendly implementation thereby reducing the area overhead compared to the conventional local assist circuitry.

II. Novel Low-swing Write Mechanism

A. Architecture

In our architecture we propose to replace the strobed local write receiver with 2 cross coupled inverters, NS-LWR (Figure 2), reducing the timing complexity associated with strobe signal generation of the local write receiver. The WL_WR activation signal for the WR MUX not only transfers the low swing information onto the local bit-lines but also serves the purpose of triggering the regenerative action of the 2 cross coupled inverters. The WL_WR signal is pulsed in order to isolate the nodes of the cross coupled pair from the highly capacitive bit-lines. This architecture also implements the differential VSS biasing technique, which allows the independent tuning of the GND connection of the cross coupled inverters of SRAM cells and the NS-LWR. SRAM cells and NS-LWR has connections to left and right vertical GND rails VSSL and VSSR. The data dependent bias application on VSSL & VSSR for the offset cancellation of the local write receiver also improves write-ability of the accessed SRAM cells as discussed in the II.C.

B. Operation

The low swing write drivers (Figure 3a) transfer the data input information (Din_i) as low swing signals on the global write bit-lines (GWBL) pair. Then the pulsed WL_WR signal activates the WR MUX of an activated local block transferring

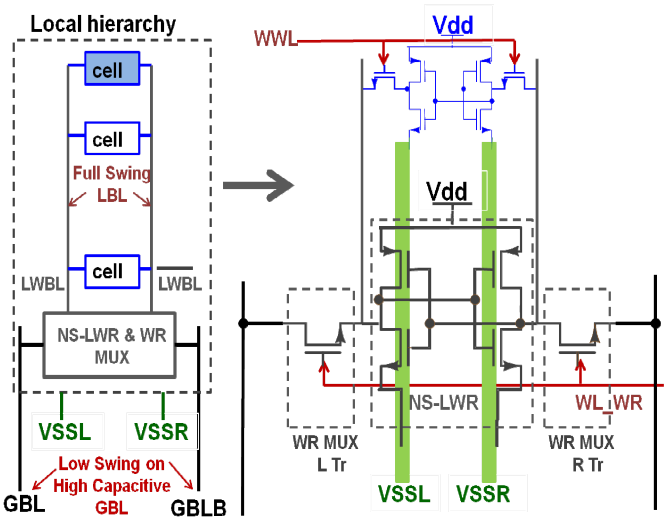


Figure 2. NS-LWR with differential VSS biasing: achieves ultra low energy WRITE operation (only write part of 8T is shown in above figure) and solves the issues related with existing LWR architecture. The NS-LWR does not impact read stability, as the WWL of our 8T cell is not activate during read.

low swing data information from the highly capacitive global write bit-lines onto the local write bit-lines pair (LWBL & LWBLbar). The regenerative action of the 2 cross coupled inverters (NS-LWR) converts this low swing data information to the full voltage level on the short local write bit-lines, so the accessed SRAM cell sees full swing bit-lines. Finally the write word line (WWL) activation signal completes the WRITE operation, flipping the internal nodes of the accessed SRAM cell. The assertion of (S, external SRAM macro pin) signal for the low voltage levels applies differential VSS bias on the GND rails of the activated matrix column (Figure 3b). The S signal and the data input signal (Din_i) applies data dependent bias on the GND line by connecting VSSL, VSSR to ±Δv. The VSSL & VSSR rails of un-activated matrix columns (CS_i = "L") are not switched and remain connected to ground. A DC-DC converter can generate the VSS bias voltage ±Δv.

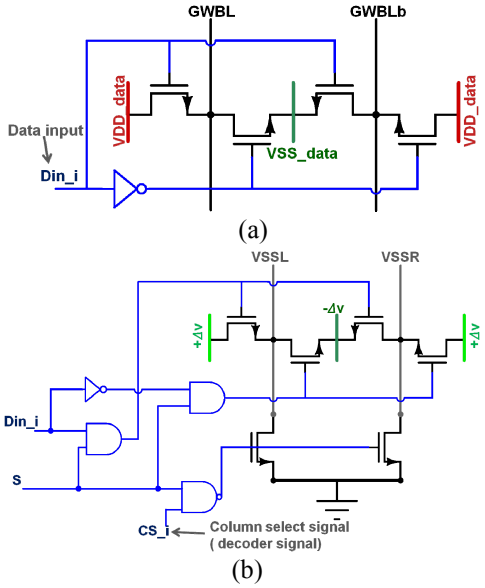


Figure 3. (a) Low swing write drivers (b) VSS biasing circuit.

C. *Variability Resilience*

Process variations can easily lead to write failures either due to sensing failure of the local write receiver or because of the degraded write-ability of the accessed SRAM cell. If MUP of the local write receiver is strong and MUPbar becomes weak (Figure 4a) due to process variations, then the risk of sensing failure increases (for writing "0"). Similarly, if MUP of SRAM cell becomes stronger and Mpass becomes weaker (Figure 4b) then the discharge of the node H becomes more difficult and the write-ability of the SRAM cell decreases.

The impact of transistor sizing in improving write-ability is effective at the high supply voltage levels but at low supply voltage levels the impact of transistor sizing is very limited in advanced CMOS technologies. Therefore, a write assist scheme is necessary to ensure SRAM cell write-ability at the scaled supply voltage levels. There are numbers of write assist techniques available to solve the degraded write-ability viz. boosted WL [10] and lowering Cell VDD [11]**.** Functional effectiveness is the most important parameter in the evaluation of the write assist technique applied. But at the same time the added power consumption and area overhead are also equally important parameters.

In this design, the differential VSS bias enabled NS-LWR provides realizes a variability resilience low energy WRITE operation (Figure 4a). The application of differential VSS bias connects VSSL to + *Δ*v, MUP becomes weak and connection of VSSR to –*Δ*v increases the strength of MUPbar. This tuning of MUP and MUPbar reduces the mismatch offset and the sensing failure is avoided. The application of differential VSS biasing of 0.1V reduces the sigma Voffset by 25%, based on the importance sampling simulations at VDD 0.55V.

The application of a differential VSS bias on NS-LWR results in a Negative BL mechanism. With the result asserted SRAM cell experiences two write assist techniques: selective VSS

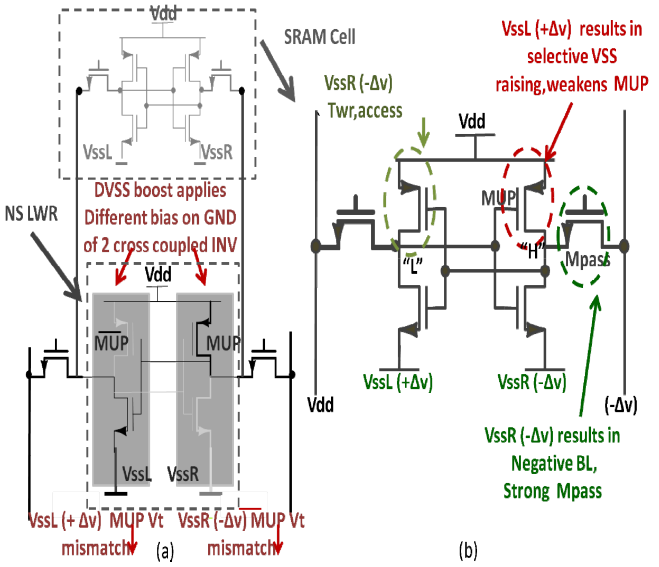


Figure 4: Differential VSS bias reduces mismatch offset of NS-LWR & improves write margin for SRAM cell.

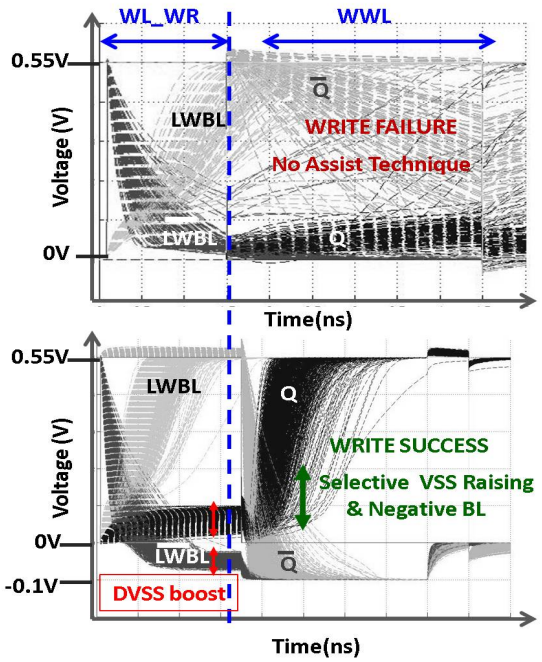


Figure 5: Variability Resilience: SRAM cell write-ability improvement.

raising (due to differential VSS biasing of VSSR and VSSL) and Negative BL (writing "0" ,Figure 4b). The positive VSS bias applied weakens MUP of "H" side of the SRAM cell thereby improving write-ability of the accessed SRAM cell. The negative VSS bias applied on the complement GND signal have 2 advantages: first it makes the rise time faster during the WRITE operation thereby improving the write access time, -0.1V of differential VSS bias results in 24% improvement for the slow NMOS and slow PMOS process corner. Secondly it pulls the bit-line below GND level (-*Δ*v) and generates the negative bit-line for the accessed SRAM cell without any extra added cost. The selective VSS raising and Negative BL mechanism increases the SRAM cell write-ability (Figure 5) and the probability of write failure for the worst corner (slow NMOS and fast PMOS) by the factor of 10^3 x at the scaled VDD levels (VDD = 0.55V).

There is no adverse impact of differential VSS bias applied on the write-ability of the SRAM cell for the opposite mismatch scenario (MUP is weak & Mpass is strong), rather it increases the speed of the WRITE operation. The application of differential VSS bias has some negative impact on the stability of the un-asserted SRAM cells of an activated column. But there is no risk of instability, worst case simulated SNM hold at VDD= 0.55V with 0.1V differential VSS biasing is greater than zero. There is 27% degradation of the mean SNM hold of the un asserted SRAM cells with the application of 0.1V differential VSS bias at VDD = 0.55V.

III. LOCAL READ BUFFER

The access speed is dominantly dependent on the rate at which the accessed SRAM cell discharges the bit-line. The Iread is severely impacted by the increased process variations with technology scaling. The access speed is particularly problematic for low supply voltages when (VDD-Vt) is small.

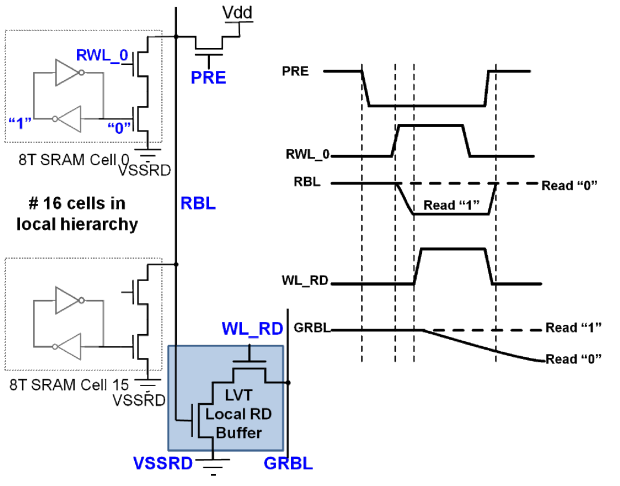
Up scaling the transistor sizes provide a linear improvement in nominal Iread, but this improvement is largely offset by the larger bit-line capacitances

In the conventional high performance SRAMs, memory cells are placed in local hierarchy with connection to short local bit-lines (Figure 1).The small sized SRAM cell has to discharge only a small capacitance. The RBL swing is then transferred onto global bit-lines through a static or dynamic local read buffer [3], [12]. The presence of the logic gates not only increases area but also results in complex memory matrix optimization. Secondly the conventional high performance SRAMs utilizes low Vt transistors thereby increasing the static energy consumption.

This design utilizes high Vt transistors for the 8T SRAM cell and proposes a local assist circuitry consisting of 2 upsized LVT NMOS transistors based read buffer (Figure 6a). The sources of 8T SRAM cells read buffers and local read buffer (upsized LVT transistor) are both connected to VSSRD. In other words both SRAM cell read buffer and the local LVT read buffer sink current in VSSRD, which is kept floating for all non-accessed matrix columns. The high Vt 8T SRAM cell, floating VSSRD and low swing pre charge voltage for GRBL minimizes the leakage power. The leakage current is reduced by 40x for the worst case (fast NMOS and fast PMOS process corner). The local read buffer delivers more current compared to the minimum sized HVT read buffer of an accessed SRAM cell and improves the access speed.

The upsized 2 stack LVT NMOS transistor along with the write assist circuitry used during WRITE offers higher degree of flexibility in memory matrix optimization as discussed in section IV, resulting in area reduction compared to conventional assist circuits.

The matrix column for an accessed SRAM cell is activated by connecting its VSSRD port to the GND. The asserted 8T SRAM cell discharges the local read bit-line depending on the stored data information (Figure 6b). Then the local read buffer is activated by WL_RD signal.



(a) (b)

Figure 6: Local read buffer (structure similar to 2 stack NMOS transistors of 8T SRAM cell) in local hierarchy is shared for 16 8T SRAM cells (read part shown in above figure).

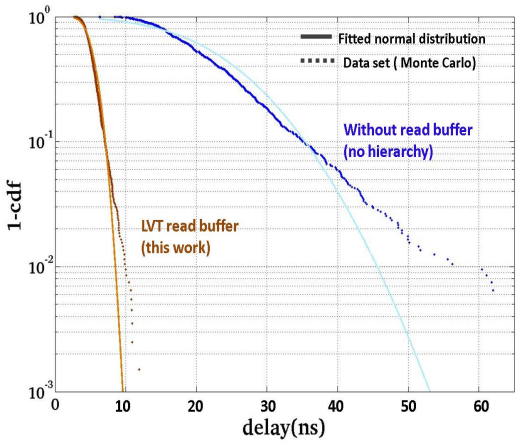


Figure 7: Distribution of the delay (read access) from RWL activation to 100mV swing on GRBL for 1K Monte Carlo runs at VDD = 0.55V for the worst case (slow NMOS and slow PMOS process corner).

Local read buffer transfers the information from the local read bit-line to GRBL to be sensed by the global sense amplifiers. The insertion of a local read buffer (2 upsized LVT transistors) in the local hierarchy mitigates the impact of the small cell read current (due to the use of high Vt 8T SRAM cells) on the memory access speed. The presence of local read buffer results in overwhelming improvement in variability resilience at scaled voltage levels (Figure 7). The nominal delay with local read buffer is 5.35 ns at VDD = 0.55V compared to nominal delay of 22.9 ns at VDD=0.55V.

IV. Local Assist Circuit Layout

The physical regularity of SRAM layout enables the use of litho optimized specialized DRC rules. The advantage of ultra regular layout of SRAM matrix in achieving area reduction is quite obvious. However, achieving the same benefit from the logic circuit is difficult because the logic circuit layout tends to be irregular. As a result the conventional logic circuit based local assist techniques [3][6-8][12]complicate the litho optimization of the memory matrix.

The local assist circuitry as proposed in this work consisting of NS-LWR, WR MUX and local read buffers are easy to map onto regular design fabric, similar to SRAM cells. The components of the local assist circuitry consist of 2 cross coupled inverters of the local write receiver, 2 NMOS pass transistor of the WR MUX and the 2 stack NMOS transistor of the local read buffer resemble an 8T SRAM cell.

The additional NMOS pre charge transistor for the local read bit line is implemented in the local read buffer region (Figure 8). In other words this local assist circuitry facilitates shape–level regularity requirement to take advantage from the litho optimization. Enforcing shape–level regularity for litho optimization is a difficult task with the existing conventional local assist techniques. Therefore the 8T SRAM cell type implementation of the proposed local assist circuitry offer enhanced flexibility for embedding the logic circuit into the memory matrix at a reduced area cost.

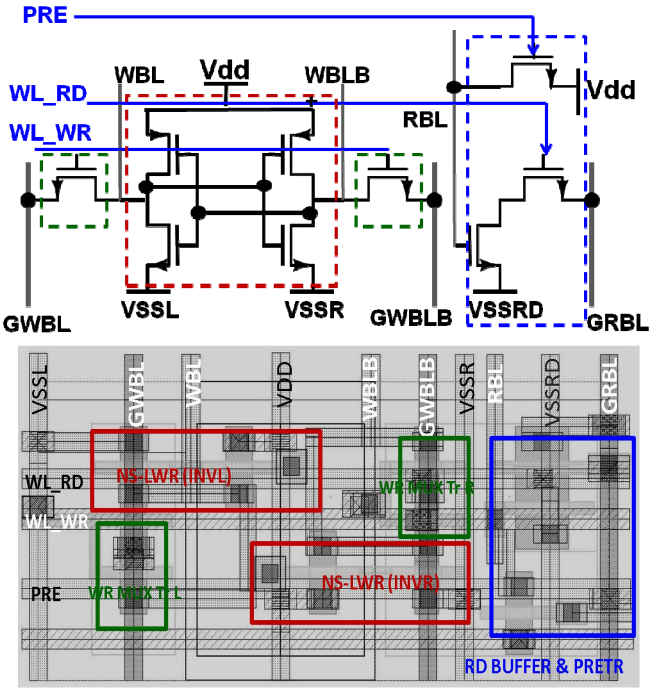


Figure 8: 8T SRAM cell type layout of the local assist circuitry: ~2x of actual 8T SRAM cell.

V. Simulation Results

A. *Write Margin Improvement of SRAM Cell*

The differential VSS bias applied to NS-LWR generates negative bit-line in addition to selective VSS raising thereby improving the write ability. The data dependent ± 0.1V differential VSS bias results in 2.5x improvement in the write trip point (DC simulations) for the worst corner (slow NMOS and fast PMOS) (Figure 9).

B. *Energy Consumption*

The energy consumption (Figure 10) is 10x less compared to the conventional full swing bit-lines used for 8T SRAM. In conventional designs highly capacitive non hierarchical bit-lines with all the SRAM cells connected (512 cells) are switched full swing whereas in this work full swing voltage is used only for the local bit-lines connected to 16 cells and there are 32 such local blocks connected to low swing highly capacitive global bit-lines. NS-LWR used in local hierarchy for the amplification of low swing data input information reduces the timing complexity associated with the existing state of the art LWR [6-8]. Secondly the litho friendly SRAM cell type layout enables compact layout, thereby reducing bit-line wire capacitances. This directly maps into reduction in energy consumption. The energy consumption per bit of this design for the column height of 512 cells with 32 local blocks (16 cells per local block) is 40% less compared to the existing state of the art similar sized LWR [6-8] for the worst process corner (fast NMOS and fast PMOS).

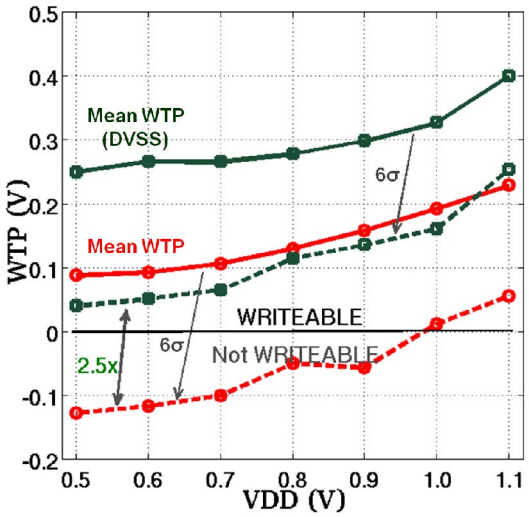


Figure 9. Write trip point (WTP) improvement with DVSS assist: Negative BL and VSS raising.

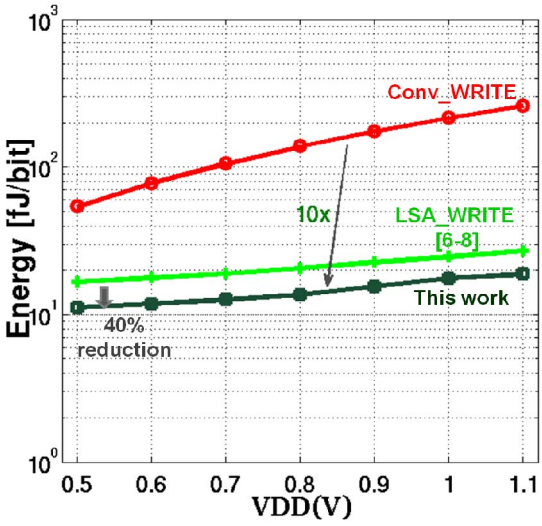


Figure 10. Energy Consumption per bit for coulmn height of 512 cells.

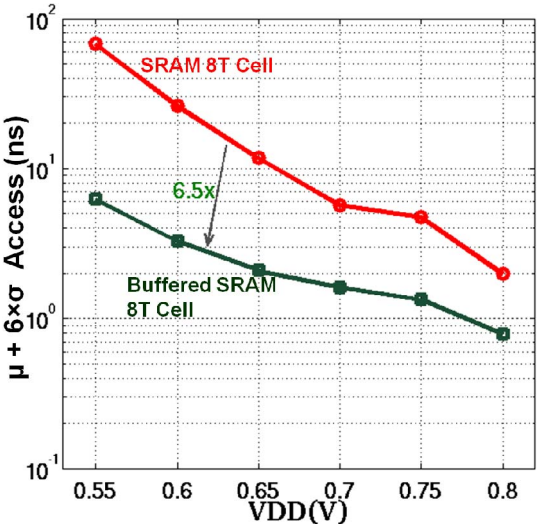


Figure 11. READ T_{Access} reduction with local upsized READ buffer.

C. *Access Time for READ Operation*

Upsized LVT read buffer in the local hierarchy delivers more current compared to the high Vt 8T SRAM cell and achieves 6.5x improvement in READ access speed for the worst process corner (slow NMOS and slow PMOS) (Figure 11).

D. *Area Reduction*

(Figure. 12) shows the best effort layout of the strobed LWR and of the non strobed LWR integrated in a local bit slice with 16 SRAM cells per local bit-line. This work utilizes logic DRC rules based SRAM cells due to the non –availability of litho-optimized parametric SRAM cells for academic purposes. The area overhead of proposed solution is only 9% compared to 38% with the existing solutions [6-8]. Firstly, the non strobed LWR reduces the transistor count compared to the conventional strobed LWR [6-8]. Secondly the differential VSS bias applied for the offset mitigation further relaxes the transistor sizing requirement compared to the conventional LWR. Thirdly the SRAM cell type structure of NS-LWR and associated WR MUX enables compact pitch matched layout. With the result area of our local assist circuitry is 31% less compared with the conventional local assist circuitry.

VI. CONCLUSION.

The local assist circuit techniques proposed here addresses the issues of increased mismatch offset and degraded write-ability associated with the increased device variations at the scaled voltage levels for the advance sub-nanometric technologies and achieves an ultra low energy operation. NS-LWR reduces the transistor count and timing complexity associated with the conventional strobed LWR. The differential VSS bias application mitigates the impact of mismatch offset, therefore the probability of sensing failure is much reduced. Reduced timing complexity and transistor sizes reduce the energy consumption of NS-LWR compared to the conventional LWR. The WRITE energy consumption is 10x compared to the conventional full swing bit-lines (full swing switching on highly capacitive bit-lines, column height of 512 cells) and 40% less compared to the existing state of the art techniques [6-8]. The differential VSS bias applied on NS-LWR results in negative bit-line on the VSS side of local bit-line and selective VSS raising. Therefore the actual cell to be written will experience two write margin improvement techniques VSS raising & Negative BL at the only cost of differential VSS bias applied to the NS-LWR. The upsized LVT local read buffer reduces the READ access time by 6.5x and the use of low swing GRBL and floating VSSRD reduces the leakage. The area overhead of this solution is only 9%. The physical regularity in the layout of the local assist circuitry permits the litho optimization thereby eliminating the memory matrix sub array design complexity associated with the placement of logic circuits [3,6-8,12]. Thus the proposed circuit techniques offer a better area-energy-performance trade-off.

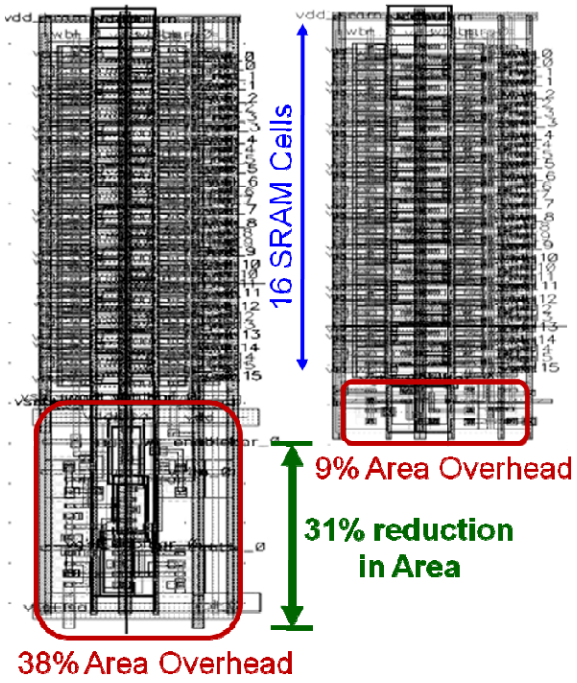


Figure 12. Area reduction.

REFERENCES

[1] L.Chang, et al., "Stable SRAM Cell Design for the 32nm Node and Beyond," *Proc. of IEEE Symposium on VLSI Technology*, 2005, pp. 128-129.

[2] Y.Hiroyuki, Embedded SRAM Trend in Nano-Scale CMOS, *Memory Technology, Design and Test*, 2007, pp19–22.

[3] L.Chang, et al.,"An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches", *IEEE J.Solid State Circuits* , vol 43, no.4, pp. 956-963, April 2008.

[4] K.Mai, et al., "Low-power SRAM design using half-swing pulse-mode techniques," *IEEE J.Solid State Circuits*, vol.33, no.11,pp.1659-1671, November 1998.

[5] K.Kanda, et al., "90% Write Power-Saving SRAM Using Sense-Amplifying Memory Cell," *IEEE J. Solid State Circuits,* vol.39, no.6, pp. 927-933, June 2004.

[6] B.D. Yang, et al., "A Low-Power SRAM Using Hierarchical Bit Line and Local Sense Amplifiers," *IEEE J.Solid-State Circuits*, vol.40, no.6, pp. 1366-1376, June 2005.

[7] S.Cosemans, et al., "A Low –Power Embedded SRAM for Wireless Applications", *IEEE J.Solid State Circuits*, vol.42, no.7, pp 1607-1617, July 2007.

[8] V.Sharma, et al.,"A 4.4pJ/Access 80MHz, 2K Word X 64b Memory with Write Masking Feature and Variability Resilient Multi-Sized Sense Amplifier Redundancy for W.S.Nodes", *Proc. IEEE ESSCIRC*, pp 358-361, 2010.

[9] M.J.M. Pelgrom, et al., "Matching properties of MOS transistors," *IEEE J.Solid-State Circuits*, vol.24, no.5, pp. 1433-1439, Oct 1989.

[10] Y.Morito, et al., "An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design under DVS Environment", *Proc. IEEE Symposium. VLSI Circuits*, 2007, pp. 256-257.

[11] K.Zhang, et al., "A 3-GHz 70-Mb SRAM in 65nm CMOS Technology with Integrated Column-Based Dynamic Power Supply," *IEEE J.Solid-State Circuits,* vol. 41, pp. 146-151, 2006.

[12] S.Ishikura, et al., " A 45 nm 2-port 8T SRAM Using Hierarchical Replica Bitline Technique With Immunity From Simultaneous R/W Access Issues," *IEEE J.Solid-State Circuits*, vol 43, pp 938-945, 2008.