Dynamic-Threshold CMOS SRAM Cells for Fast, Portable Applications

Azeez J Bhavnagarwala, Ashok Kapoor[‡] and James D Meindl

Microelectronics Research Cntr. and the School of Elec. and Comp. Eng., Georgia Inst. of Tech., Atlanta GA 30332

[‡] LSI Logic Corporation, Milpitas, CA 95035

Abstract[†]

A novel quad-rail CMOS SRAM cell architecture that doubles cell read current, improves cell static noise margin (SNM) by 70%, increases cell immunity to SER and lowers cell standby power by over an order of magnitude is proposed. These improvements are achieved by implementing a scheme of WL transition triggered pulses on source and substrate terminals of cell inverter transistors that share a common WL.

Introduction

With scaling of MOSFET dimensions, microscopic variations [1-3] in number and location of dopant atoms in the channel region of the device induce increasingly limiting electrical deviations in device threshold voltage (Fig. 1) [3]. These atomic level intrinsic fluctuations cannot be eliminated by external control of the manufacturing process and are most pronounced in minimum geometry transistors commonly used in area constrained circuits such as SRAM cells [4]. Narrow width effects, SER, low voltage operation, temperature and process variations, and parasitic transistor resistance all contribute additionally to the increasing instability of a conventional 6T SRAM cell as well [5-7]. With projected increases in percentage of chip transistors devoted to cache [8], in high performance microprocessors (Fig. 2) and ASICs, subthreshold leakage currents from an overwhelming number of inactive cells are projected to become larger than the AC currents from a much smaller number of active circuits [9] placing limits on the scaling of threshold voltage of cell transistors. The Bit Line (BL) delay determined primarily by BL capacitance, cell read current and sensitivity of the sense amplifiers, increasingly limits SRAM performance [10] since BL capacitance and sense amplifier sensitivity do not scale as rapidly as minimum feature size [5]. Raising the V_T of cell transistors reduces standby power dissipation from the cell array to tolerable levels but imposes a severe penalty on BL delay by compromising cell read current drive of minimum geometry cell transistors. Several innovative SRAM cell architectures [11-13] have been proposed to alleviate one or more of the above limitations facing SRAM cell scaling. However, in all of these (i) the improvements in cell drive obtainable by cell voltage/overdrive boosting are dampened by the high V_T of cell transistors and (ii) the opportunity to use the substrate terminal of the cell transistors to control leakage and improve cell stability and performance is not pursued. In this work, we propose a new circuit technique that *does not require trading off subthreshold leakage for cell drive*. By dynamically biasing the source-body junctions of all cell transistors, in cells driven by a common WL, improvements achieved in cell read current by boosting cell voltages are not impeded by high V_T required for low leakage. Off-chip quad-rail supplies are required and are shown to impose a minimal cost in cell area, wiring complexity and processing.

Speed improvements by pulsing source terminals

Driving the *source terminal of cell inverter MOSFETs* (Fig. 3 a, b) marginally outside the V_{dd}-Gnd range, by voltages less than the built-in potential of a p-n junction (to avoid excessive forward bias leakage) translates into substantial increases in read current. This increase in read current is larger than that obtained by driving any of the other MOSFET terminals by the same voltage swing because all of the following contribute to drive increase (i) increase in gate-source voltage (ii) increase in drain-source voltage for devices with finite output resistance and (iii) reductions in device threshold voltage due to forward bias on the source-body p-n junction. The series-connected cell access transistors drive more current as well since the low and high storage nodes of the cell are driven past the V_{ss} and V_{dd} voltage levels (Fig. 4b), boosting the gate-source, drain-source and source-body voltages of the access transistors. *Both - the BL discharge path as well as the BLB charge path are dilated* translating into a doubling of cell read current (Fig. 4a).

Stability and BL power improvements

In a conventional CMOS SRAM cell, the '0' storage node voltage, rises above ground during a read access due to voltage division along the access and inverter pull-down NFET devices between the precharged high bit line and the ground terminal of the cell. The ratio of the widths of the pull-down transistor to the access transistor, commonly referred to as the *cell ratio* or *beta ratio*, determines how high the '0' storage node rises during a read access. Smaller cell ratios translate into a bigger voltage drop across the pull-down transistor requiring a smaller noise voltage at the '0' node to trip the cell. Driving the source terminals of the cell pull-down NFETs below V_{ss} ensures enough room for the voltage drop across the pull-down device so that

[†] This work was supported by the Defense Advanced Research Project Agency (Contract: F3361595C1623), the Semiconductor Research Corporation (SJ-374-002) and by LSI Logic Corporation

0-7803-6598-4/00/$10.00 © 2000 IEEE

the '0' storage node does not rise above ground during a read access (Fig. 4b). An alpha particle hit thus requires more noise charge to drive the storage node through a larger swing to flip the state of the cell. *This directly translates into a higher immunity to SER.* In a conventional CMOS SRAM cell, lowering the BL precharge voltage lowers the '1' storage node voltage during a read access, due to a voltage divider between the precharged BL and the source terminal of the pull-up PFETs. The '1' storage voltage deteriorates very rapidly with lower BL precharge voltages due to the higher on-resistance of the pull-up PFETs compared to the pull-down NFETs, making it necessary in conventional CMOS SRAM cells and for the cells reported in [11-13] to precharge BLs high. *High BL precharge voltages restrict potential opportunities to improve BL delay by driving the BL pair in opposite directions during a read access. High BL precharge voltages also require more energy to recover the BL voltage after a write operation or after extended periods of cell inactivity when clocked BL loads are used.* Clocked (AC) BL loads are typically used in high performance SRAMs so that the minimum geometry cell transistors do not have to fight DC BL loads when pulling the BL down during a read access. For source-pulsed DT SRAMs, pulsing the source terminals of the pull-up PFETs ensures enough room for voltage drop across the pull-up PFETs when the BL pair is precharged to a low, non-zero voltage ($V_{dd}/3$). The sum of the BL charge and discharge currents (effective read current) of the source-pulsed DT SRAM cell, are double that seen in a conventional cell with identical cell transistor geometries (Fig. 4a). Since the SL and the PL are triggered by V_{dd}-V_{ss} transitions of the WL, unconventional drivers are required to sense the WL transitions and drive the SL and PL to voltages outside the V_{dd}-V_{ss} range. The schematics of these drivers are shown in Fig. 5 a, b and the waveforms (from HSPICE) they generate are shown in Fig 6.

Dynamic V_T cell transistors and subthreshold leakage improvements

Subthreshold leakage from unaccessed cells in large arrays is lowered substantially by reverse biasing the source-substrate junctions of all cell transistors when the WL of the cells is deselected. Thresholds are lowered to their nominal values by driving the substrate terminals to V_{dd} and V_{ss} when the WL is selected. This technique requires a triple-well process where all of the N channel devices in cells accessed by a common WL are isolated and share a common tub. The same requirements apply to the p-channel devices as well. The substrate-source reverse bias may be as high as V_{dd} which raises the device threshold voltages by 100 mV lowering the leakage currents by over an order of magnitude (Fig 7). Channel doping profiles of the cell transistors may be chosen to yield a maximum rise in threshold voltage in response to the reverse bias on the source-substrate junction. Special drivers (Fig. 8a, b) triggered by WL transitions are required again to drive the substrate terminals of cell transistors outside the V_{dd} – V_{ss} range (Fig 9). These drivers may require additional processing since they are driven by larger voltages than those seen by cell transistors. The substrate terminal drivers are loaded by junction capacitance of each terminal, which is smaller than the two gate-input capacitance seen by the WL driver. This permits the substrate to respond at least as fast as the WL causing the device thresholds to be lowered before or at about the same time that the WL signal rises.

Overhead in delay, power, wiring and processing

Since the SL, PL, SubP and SubN lines are driven by circuitry after the arrival of the WL signal, the additional delays introduced reduce the gains obtained by doubling the cell read current. The BL response (Fig. 10) shows an improvement in WL-BL delay of 33% in the SP_DT SRAM cell over a conventional cell with identical cell transistor geometries the above overhead in delay notwithstanding. Additional dynamic power is dissipated in driving the SL, PL, SubN and SubP lines. Some of this is recovered by the energy efficient low BL precharge voltage and the overheads may be kept at a minimum with the commonly used hierarchically divided WL architecture, that employs short WLs. Since the SP_DT cells do not require V_{dd} and Gnd rails, the overhead in cell area comes from only two sources: (i) additional area consumed by separate source and substrate contacts and (ii) requirements of a twin-tub process impose an additional overhead on cell area as well. However, since the SP_DT cell with minimum geometry transistors is substantially more stable than a conventional SRAM cell, this overhead in cell area may be kept at a minimum by using unity cell ratios. The 1999 Int'l Roadmap for Semiconductors (ITRS) projects multiple supply voltage requirements for analog and RF blocks in a heterogeneous system-on-a-chip making the availability of boosted power supplies for SP_DT cells a lesser concern.

Summary and Conclusions

A novel CMOS SRAM cell architecture that improves stability by over 70%, WL_BL delay by 33%, BL power by an order of magnitude and subthreshold leakage by over an order of magnitude is reported. The cell requires a marginal overhead in cell area and wiring for its quad-rail power supplies. These improvements are achieved by pulsing the source and substrate terminals of the cell transistors selectively outside the V_{dd} – Gnd range. This new architecture considers holistically all of the challenges facing conventional

scaled CMOS SRAMs of increased leakage, slower BL response and increasing instabilities.

References

[1] R. W. Keyes, "The Effect of Randomness in the Distribution of Impurity Atoms on FET Threshold," *App. Phys.*, 8, pp. 251-259, 1975.

[2] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental Study of Threshold Voltage Fluctuations using an 8K MOSFET Array," *Symp. VLSI Tech.*, pp. 41-42, Jun. 1993.

[3] J.D. Meindl et al, "Impact of Stochastic Dopant and Interconnect Distributions on Gigascale Integration," *Proceedings of the 1997 IEEE ISSCC*, pp.232-233, February 1997.

[4] K. Itoh, "Low Power Memory Dsgn.," *Tutorial, 1997 ISLPED*

[5] B. Bateman, "High Performance SRAM Design," *Tutorial, ISSCC*, Feb. 1998.

[6] E. Seevinck, F. List, and J. Lohstroh, "Static-Noise Margin Analysis of MOS SRAM Cells," *IEEE JSSC*, Vol. SC-22, No. 5, pp. 748-754, Oct. 1987.

[7] M. Inohara et al, "Highly Scalable and Fully Logic Compatible SRAM Cell Technology with Metal Damascene Process and W Local Interconnect," *IEEE Symp. VLSI Tech.*, Jun. 1998, pp. 64-65.

[8] D. Burger, "System-Level Implications of Processor-Memory Integration," *24th Int'l Symp. on Comp. Arch.*, June '97, pp.

[9] K. Itoh, "Trends in Low Power RAM Circuit Technologies," *Proc. IEEE*, Vol. 83, No. 4, Apr '95, pp. 524-543.

[10] M. Ishida et al, "A Novel 6T-SRAM Cell Technology Designed with Rectangular Patterns Scalable beyond 0.18mm Generation and Desirable for Ultra High Speed Operation," *IEDM Dig. of Tech. papers*, Dec. 1998, pp. 201-204.

[11] K Itoh et al, "A Deep Sub-V, Single Power-Supply SRAM Cell with Multi-VT, Boosted Storage Node and Dynamic Load," *IEEE Symp. VLSI Ckts.*, Jun 1996, pp. 132-133

[12] H. Mizuno et al, "Driving Source Line (DSL) Cell Architecture for Sub 1-V High Speed Low Power Applications," *1995 Symp. VLSI Ckts.*, pp. 25-26.

[13] H Yamauchi, et al, "A 0.8V/100MHz/sub-5mW Operated Mega-bit SRAM Cell Architecture with Charge Recycle Offset-Source Driving (OSD) Scheme," *1996 Symp. VLSI Ckts.*, pp. 126-127.

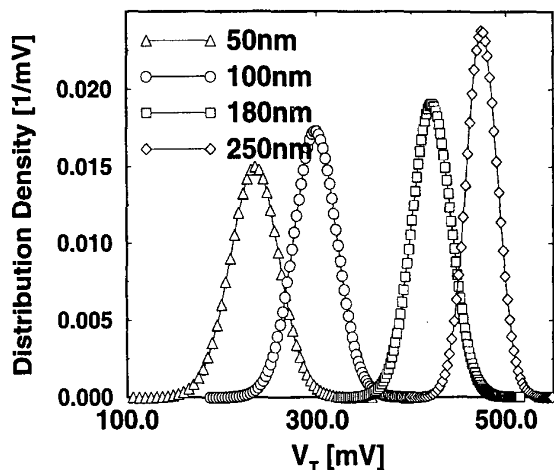


Fig. 1 : Threshold voltage distribution functions due to intrinsic fluctuations alone for representative 1997 NTRS generations [3]. Minimum geometry cell transistors become increasingly vulnerable to intrinsic fluctuations

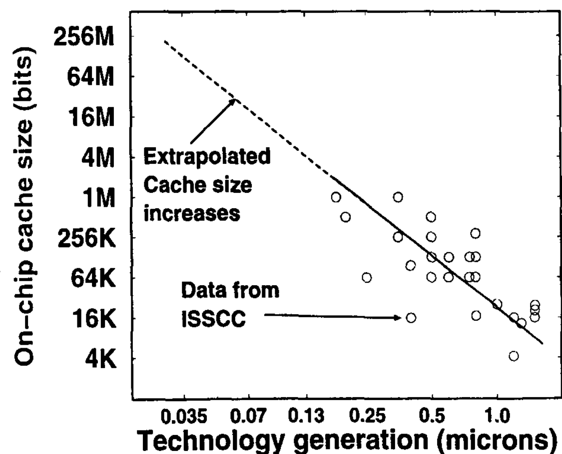


Fig. 2 : Projected increases in SRAM cache size of microprocessors by extrapolating trends over the last 5-10 years.

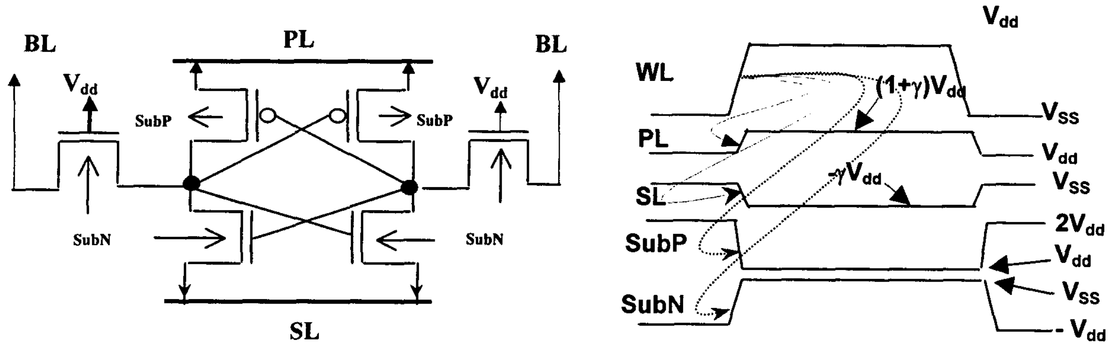


Fig. 3 a, b : Circuit schematic and timing diagram of a source-pulsed CMOS SRAM cell. Power Line (PL) and Source Line (SL) pulses are triggered by the WL signal for the entire row of cells driven by the WL during a read access. $\gamma V_{dd} < 600mV$. Substrate terminals of P and N channel devices are at V_{dd} and V_{ss} respectively during a cell access.

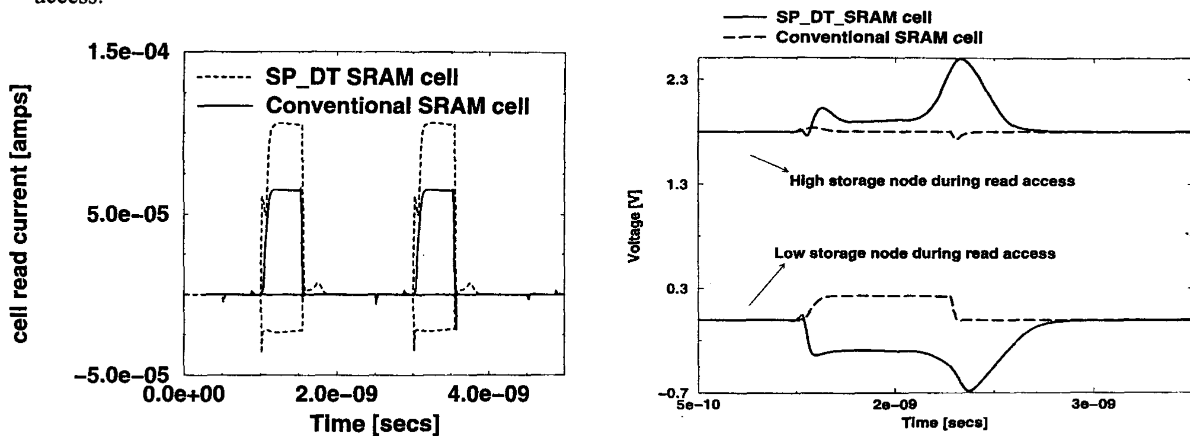


Fig. 4a, b : *HSPICE simulations of a 0.18 µm cell.* During a read access, pulsing the source terminals of the cell inverter MOSFETs drives the cell storage nodes in opposite directions past V_{dd} and V_{ss} voltage levels increasing the drive current of cell inverter transistors as well as the cell access transistors. Increases in cell drive result from higher gate-source voltage, higher drain-source voltages and forward-biased source-substrate junctions. The boosted source terminal voltages are within limits set by forward-bias leakage (600mV) and by device reliability constraints. Read current from an SP_DT cell increases unevenly because the source and substrate terminals do not trigger at exactly the same time.

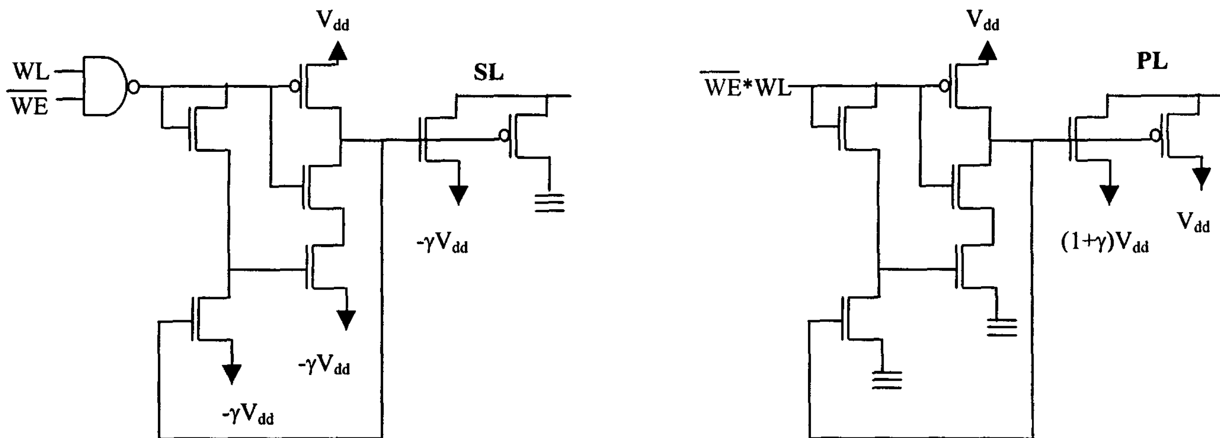


Fig. 5 a, b : Circuit schematics of Source Line and Power Line drivers. Drivers pulse the SL and PL only during a read access. Devices used in drivers may need additional processing to permit higher gate-source and drain-source voltages.

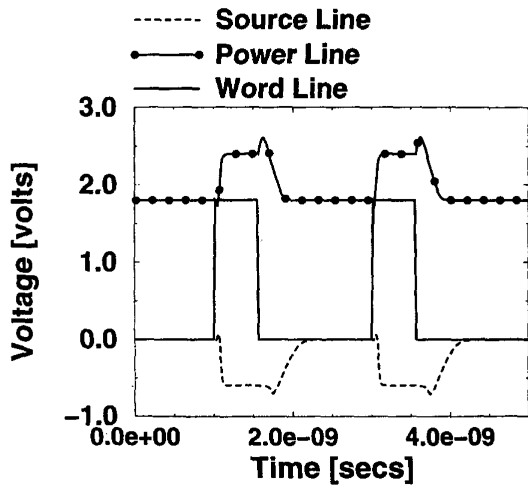


Fig. 6 : HSPICE simulations of PL and SL waveforms triggered by WL transitions

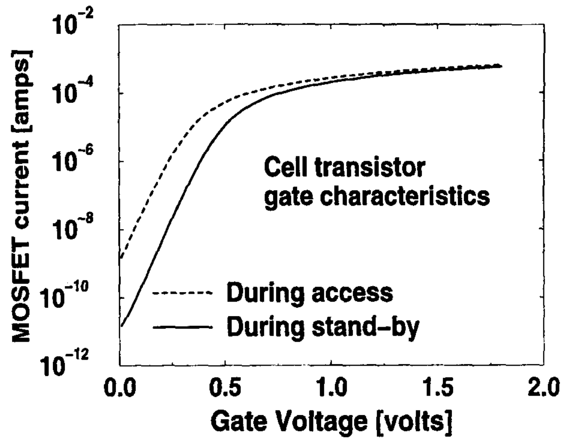
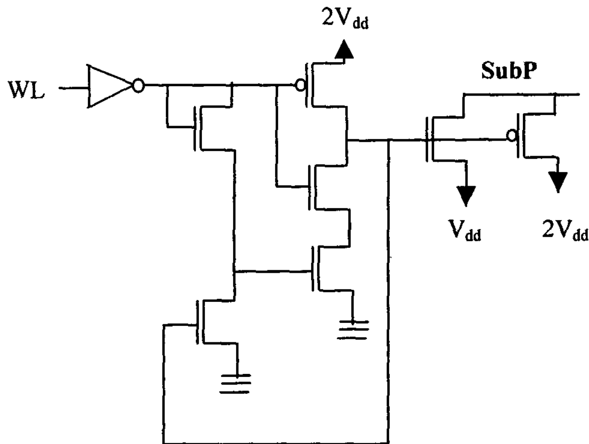


Fig. 7 : Subthreshold current reductions by dynamically reverse biasing source-body junctions by V_{dd}



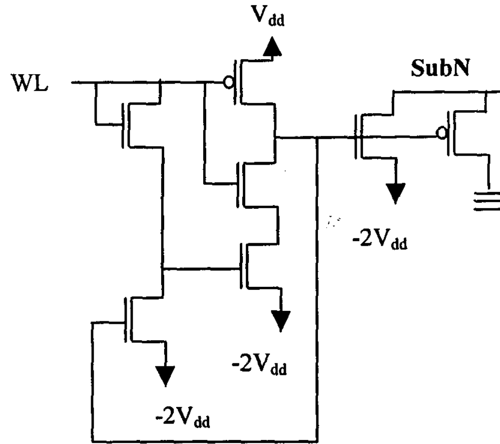


Fig. 8 a, b : Circuit schematics of PFET and NFET substrate terminal drivers. Drivers pulse the subP and subN terminals of all cells accessed by a common WL, only during a read/write access.

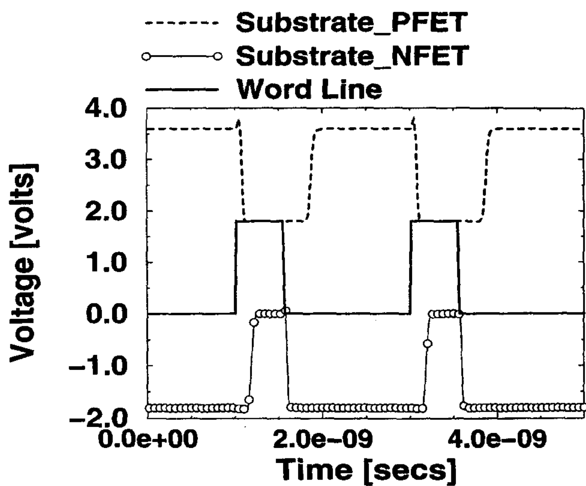


Fig. 9 : HSPICE simulations of waveforms at the substrate terminals of the PFETs and NFETs triggered by WL transitions for SP_DT SRAM cells.

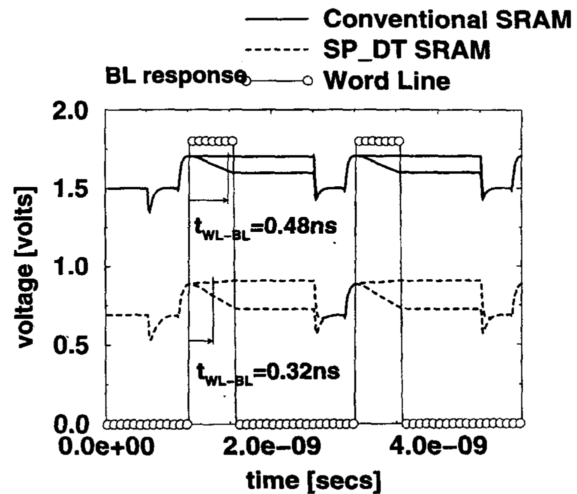


Fig. 10 : Comparison of BL response in SP_DT SRAM with a conventional 6T SRAM cell. WL-BL delay reduces 33% with identical cell transistor geometries for both cases