CIRCUIT STYLES AND STRATEGIES FOR CMOS VLSI DESIGN ON SOI

Fari Assaderaghi
IBM Semiconductor Research and Development Center
1580 Route 52, Hopewell Junction, NY 12533
Phone: (914) 892-3969
Email: fari@us.ibm.com

ABSTRACT

This paper reviews specific circuit styles and strategies employed in the design of CMOS VLSI on partially-depleted (PD) SOI. These strategies address issues and problems that arise on PD SOI circuits (mainly due to the floating-body effect) such as delay hysteresis, noise margin reduction, etc. These circuit approaches also try to utilize SOI-specific properties to achieve a larger performance gain than that of a simple re-map of a bulk design to SOI. Although many aspects of CMOS design pertaining to SOI will be covered, the emphasis will be on dynamic and static circuits and high-performance SRAM's.

1. INTRODUCTION

Applications of SOI technology in the past were limited to space and military programs where radiation-hardness was more important than performance and cost. These products were built on partially-depleted SOI while providing fixed potentials for the body of MOSFET's through well and substrate contacts, as in bulk CMOS technology. It was only recently that a radical approach was suggested on SOI, *i.e.*, use of PD SOI with *floating body nodes* [1-2]. It was demonstrated that leaving the body of SOI MOSFET's floating would not only lead to a better density, it would actually improve the circuit speed as well [2-3]. Obviously, a floating-body node could lead to some circuit issues and problems. Initially it was thought that these effects were so severe that they would impact SOI circuit functionality, and render its design unmanageable [4]. However, several successful implementations of bulk SRAM's and microprocessors on SOI with little or no change of the original designs proved otherwise [5-6]. Indeed, recently several high-performance RISC microprocessors were announced on PD SOI where all design issues were addressed and the microprocessors were built to commercial product specifications (including stress, reliability, burn-in, yield, etc.) [7-9]. These efforts not only demonstrated that PD SOI circuit issues could be handled with relative ease, they also showed examples where SOI unique features were exploited for a larger performance gain than achievable through a simple bulk-to-SOI re-map [7-9].

Here, we first describe SOI physical properties that impact design of CMOS VLSI circuits. Next, we review specific circuits that require modifications from their bulk designs due to these impacts. Suggested solutions and modifications are outlined next. These solutions include circuit changes as well as technology and process alterations affecting device behavior. The importance of a physical and predictive compact model (SPICE-like) for simulation of SOI circuits can not be overemphasized. Here, due to SOI-specific effects the need for an accurate model is even greater than in bulk design. To predict floating-body effects over long periods properly, the model needs to conserve charge as well. Otherwise, erroneous body voltages will be simulated. Based on these requirements, IBM's internal ASX model was derived from UC Berkeley's BSIM3 model [10] and was used for all of the analysis that follows.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ISLPED99, San Diego, CA, USA
©1999 ACM 1-58113-133-X/99/0008..$5.00

2. SOI CIRCUIT ISSUES AND SOLUTIONS

SOI circuit issues stem from three unique SOI features:

2.1 Film Thickness Limitation

The physical thickness of active silicon layer is limited to typically less than 200nm. This has impact on several structures: Vertical diodes used for ESD protection or in band-gap reference circuits now have to be made as lateral poly-bounded diodes, which are less area efficient. An innovative way to create more efficient diodes for ESD protection is to use dynamic threshold body- and gate-coupled diodes in SOI [11]. In these structures, MOSFET's are used in diode configurations by tying the gate, body and drain nodes all together as one terminal of the diode, and the source as the other terminal. In this structure, the total current consists of the regular body-to-source diode current, *plus* the MOS current (which is enhanced by V_t-lowering), *and* the parasitic NPN bipolar current. Therefore, the total diode current in ESD events will be much larger than that of an ordinary poly-bounded diode, demonstrated in Figures 1 and 2.

The limited SOI film thickness also leads to elimination of the well capacitance. This capacitance in bulk technology contributes to a large portion of V_{dd}-to-ground decoupling capacitance and has to be compensated in SOI by other decoupling capacitors such as thin-oxide capacitors occupying chip's white areas.

2.2 Buried Oxide Layer

Presence of the buried oxide layer reduces the thermal conductivity of SOI, leading to self-heating effects. These effects are generally very small in circuits with reasonable transition times, causing less than 2-3 degree temperature rise from bulk to SOI. The exceptions are the off-chip drivers, self-restore circuits in arrays, and pre-charge fighting circuits where the circuits are in high-current state for a significant portion of the cycle. Here the temperature rise could be as much as 10 degrees. Nevertheless, the impact on circuit delay stays minimal, but there might be some reliability concerns.

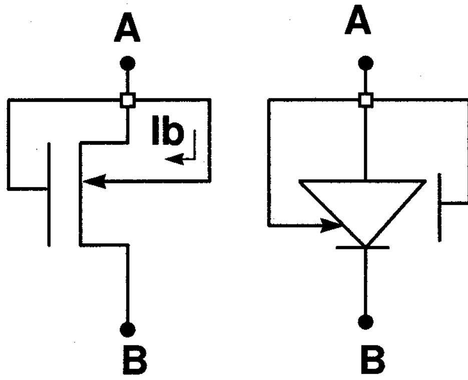

Fig. 1 **A dynamic threshold body- and gate-coupled diode is formed by tying drain, gate and body terminals of an SOI NMOSFET together.**

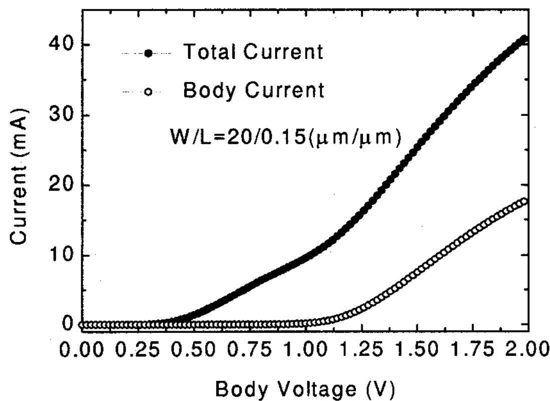

Fig. 2 **Comparison of the body-to-source diode current and the total current (with gate, drain, and body tied together).**

2.3 Floating Body Effects

The floating-body node allows the body potential to change through capacitive coupling of source, drain, and gate nodes and by impact ionization and junction leakage currents. The change in body potential leads to threshold variation of the MOSFET and possible parasitic bipolar turn-on, which have several deleterious effects on circuits. For a successful SOI circuit implementation, these effects need to be considered carefully, and proper circuit modifications have to be made. Therefore, we describe these effects and their solutions in some detail here.

In static CMOS circuits, threshold voltage change causes the device current drive to vary, leading to changes in circuit delay: Consider the simple case of an inverter. During switching events the body voltage rises and falls due to its capacitive coupling to the gate (which is typically weak due to its shielding beyond threshold), and to the drain (determined by perimeter junction capacitance). These capacitances are shown schematically in Figure 3. Figures 4 and 5 show the typical characteristics of the input and output of an inverter in a chain of inverters. The body voltage of the NFET device is shown as well. During these switching transitions, the body-voltage swings stay the same. Consequently, the high and low values of body voltage stay nearly constant, except for very small changes due to impact ionization and diode currents. Therefore, the delay hysteresis is mainly caused by different initial body voltages.

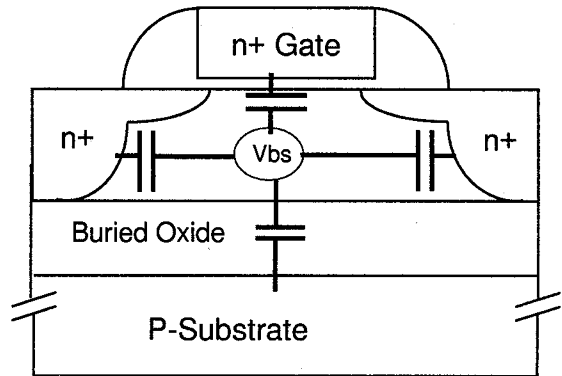

Fig. 3 **The gate capacitance, and the drain/source perimeter junction capacitances that affect the body voltage during switching.**

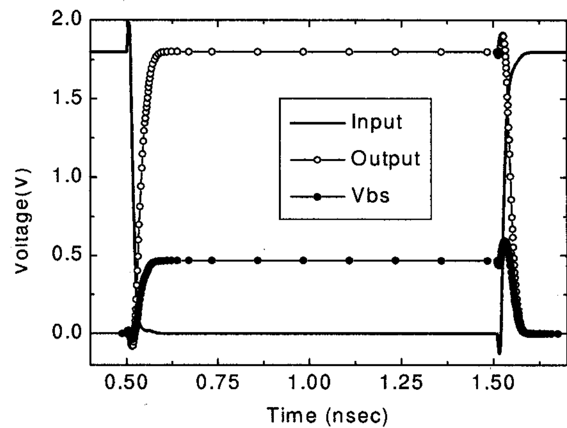

Fig. 4 **The input, output, and NFET body voltage waveforms of an inverter in a chain of inverters.**

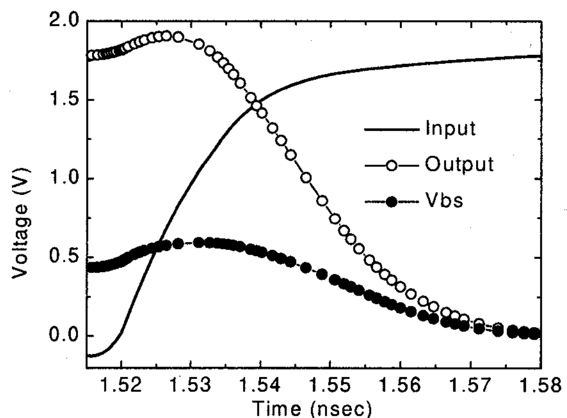

Fig. 5 **Same as Fig. 4, except details of the gate-to-body and drain-to-body capacitive coupling are shown more clearly.**

Figure 6 demonstrates different delays of an inverter in response to the rising edge of the input signal. For the "first switch" case shown as **Vin1**, the input has been low for a long time, and the NFET body voltage (**Vbs1**) is determined from diode and impact

ionization characteristics (DC behavior of the device). The value of **Vbs1** will determine the NFET device V_t just before the switching event. In the "second switch" case shown as **Vin2**, the input has been high for a long time, and the NFET body voltage (**Vbs2**) is initially at zero. As **Vin2** makes a high to low transition **Vbs2** makes a positive swing, which is entirely determined by the capacitance ratios. This value of **Vbs2** will determine the NFET device V_t just before **Vin2** makes the low to high transition. If **Vbs2** is larger than **Vbs1**, then the "second switch" delay will be shorter than the "first switch" delay (in response to the rising edge of input).

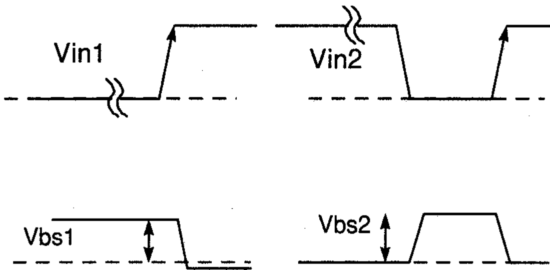

Fig. 6 **Comparison of two different inputs to an inverter, and the corresponding body voltages of the NFET device.**

Note that **Vbs1** may have temperature dependence, while **Vbs2** is effectively independent of the temperature. Therefore, varying the circuit temperature may change whether the "first switch" is faster than the "second switch" or vice versa. Also, changing the perimeter junction capacitance and gate capacitance will affect this outcome.

Stacked circuits such as NAND's, NOR's, AOI's, and OAI's in general have more complicated history effects. For these circuits the analysis can not be reduced to "first switch", "second switch", and "steady state" conditions only. The two-input NAND circuit shown in Figure 7 provides an example.

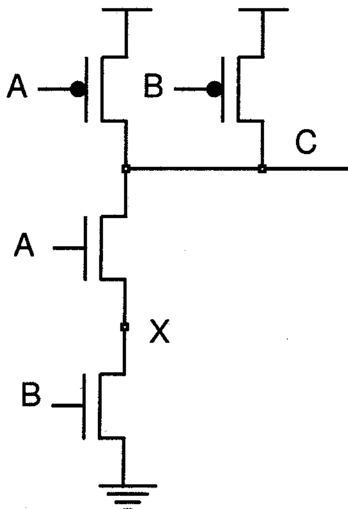

Fig. 7 **A static 2-input NAND circuit.**

Here, two different sequences of input signals are considered. In the first case, shown in Figure 8, input A is initially high, and input B is low. This allows the output node C, and the node X to be near V_{dd}. Next, input A makes a high-to-low transition, and B makes a low-to-high transition. As B transitions, node X is discharged to ground, and Vbs of transistor A temporarily has a large positive value, lowering its V_t. While **Vbs1** is high, the input A makes a low-to-high transition, causing fast switching of the output node C.

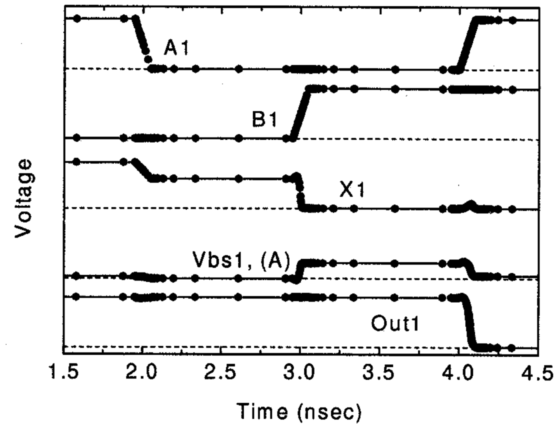

Fig. 8 **Case one of input signal sequences for the 2-input NAND circuit. Body-to-source voltage of the top NFET (Vbs1) is also shown.**

In the second case, input B is always high, and input A makes a low-to-high transition after being low for a long time. Here, the Vbs of top NFET is determined from DC conditions only, and is lower than the first case. Therefore, the switching response to input A transition will be slower.

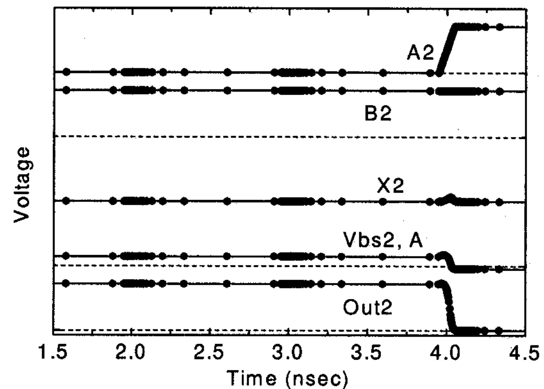

Fig. 9 **Second case of input signal sequences for the 2-input NAND circuit. The Body-to-source voltage of the top NFET (Vbs2) is also shown.**

For a direct comparison, the two scenarios are superimposed in Figure 10, and the area of interest is expanded. As seen, before input A makes the transition, Vbs of the top NFET is higher in case one than case two. This leads to a shorter delay for the first case than the second.

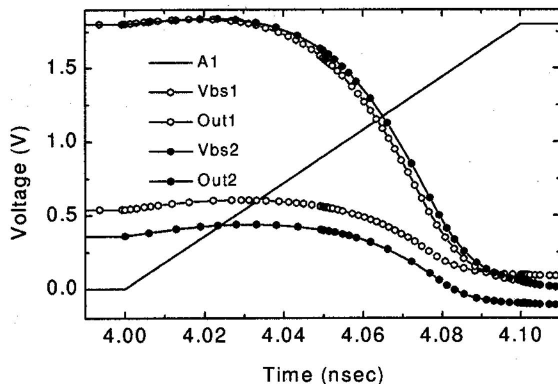

Fig. 10 **Comparison of case 1 and case 2 for input signal sequences for the 2-input NAND circuit. Body voltages of the top NFET device are also shown.**

The history effect of the delay in SOI has been studied greatly in the past, and has been mentioned as a major stumbling block of circuit design on PD SOI. However, one has to note that the largest history effect, which is the *(maximum possible delay-minimum possible delay)*, is generally not the proper variation to consider. Rather, in most designs the history effect should be defined as *(maximum delay-design point)*. Using this approach, impact of history effect on cycle time would be only 2-4%. The history-effect impact on design margin of early-mode race type circuits will be taken into account by increasing the mis-tracking assumption (already used in bulk design) by a small fraction.

Another impact of SOI floating-body node is noise margin reduction in dynamic circuits because of pass-gate leakage condition [12]. This leakage current is caused by a combination of transient V_t-lowering and bipolar turn on. Figure 11 is used to explain this transient leakage current. When the gate voltage of an SOI NFET is at zero and the source and drain are at V_{dd} for an extended time, the body voltage will eventually charge up to the source and drain voltage, by the reverse junction leakage currents. Now if source is abruptly lowered to ground, the body voltage will follow it as well due to capacitive coupling. However, the drain and gate capacitances will prevent the body from reaching the ground.

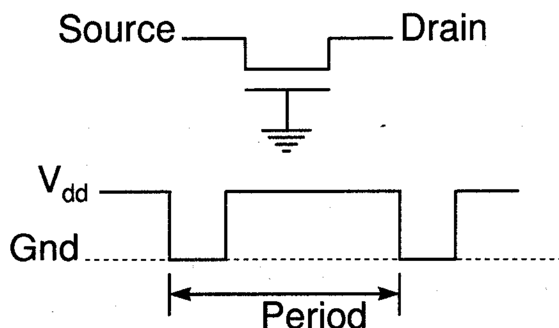

Fig. 11 **Schematic diagram for pass-gate leakage current. The drain is at V_{dd}, and the source is pulsed from V_{dd} to ground.**

Therefore, temporarily a large Vbs can exist, leading to bipolar turn on and MOS V_t lowering. Both of these effects cause a leakage current to flow between the source and the drain. Note that this leakage, although small, can be several orders of magnitude larger than typical I_{off} of the device.

Typical measured characteristics of this leakage current using a high-speed setup is shown in Figure 12 [3]. As seen, for the maximum leakage current to develop both source and drain terminals have to stay high for a long period of time. This time is required for the junction leakages to charge up the body, and it is dependent on V_{dd}, temperature, and the diode characteristics.

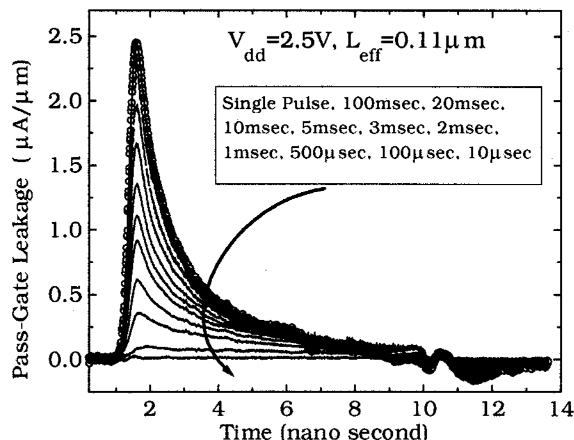

Fig. 12 **Transient pass-gate leakage current as a function of time, for different source input periods.**

Before describing circuit complications due to this leakage current and possible remedies, we describe the technology impacts: One has to note that the magnitude of the leakage current is strongly dependent on the parasitic bipolar current gain, beta. Through process and technology solutions, beta can be reduced drastically even for extremely short channel devices, as demonstrated in Fig. 13. One also should not ignore the influence of the MOSFET current. The transient leakage current that is normally bipolar dominated, can become MOS dominated if the threshold voltage of the device is made very low. Note that when source is pulsed low, the transient positive Vbs will lower the initial V_t further, and increase the subthreshold leakage current. This is demonstrated in Figure 14, which depicts simulation results for NMOS devices with different threshold voltages.

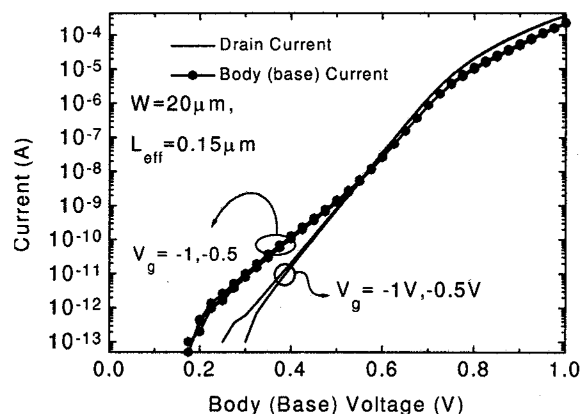

Fig. 13 **Gummel plot of the parasitic bipolar transistor. Note that for Vbs values below 0.6V, current gain is less than unity.**

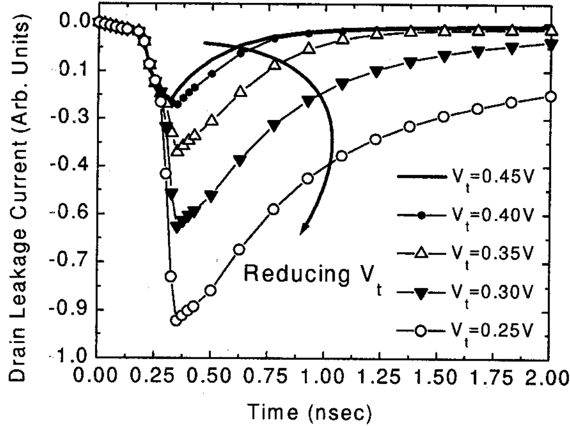

Fig. 14 **Simulation of pass-gate leakage for devices with different initial linear threshold voltages (V_t at low V_{ds}).**

The circuit in Fig. 15 is used to demonstrate the negative effect of the transient leakage current on dynamic circuits. Figures 16 and 17 show the input and output behavior. Since input A is initially high and input Z is low, both nodes X and Y will be near V_{dd}, and the bodies of B through N MOSFET's will be charged to near V_{dd} as well. During the evaluate phase, if input Z is raised and there is some input noise on B through N devices, the transient leakage currents of these devices will cause a droop in the precharge node (node X).

Under severe conditions and when the PFET keeper device is weak, the droop in the precharge node X will be too large, and the circuit will evaluate to a false value.

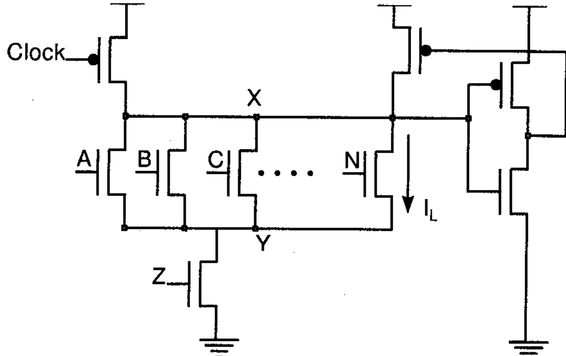

Fig. 15 **Dynamic circuit used to show the effect of transient leakage current on noise margin.**

Many circuit solutions are proposed for this problem, including pre-discharging the intermediate node, re-ordering of the NFET logic tree, re-ordering the inputs to multi-fingered stacked devices, use of conditional feedback, and re-map of the logic [7]. We demonstrate some of these solutions below through examples. In each case the original design in bulk technology and its possible variation in SOI are compared.

In bulk dynamic circuits the intermediate nodes are normally pre-charged to avoid the charge-sharing problem. This will exacerbate the pass-gate leakage current in SOI. Since in SOI the junction capacitances are much lower than in bulk CMOS, the intermediate nodes can be *pre-discharged.* This prevents the bodies of transistors high in the stack from charging to high voltage. Figures 18 and 19 demonstrate this solution where node Y is pre-discharged for SOI. Note that although the junction capacitance is significantly reduced in SOI, the gate-to source/drain overlap capacitance is still present. Therefore, this solution can still lead to some degree of charge sharing problem even for the SOI implementation. Another solution is to re-order the inputs to multi-fingered stacked devices.

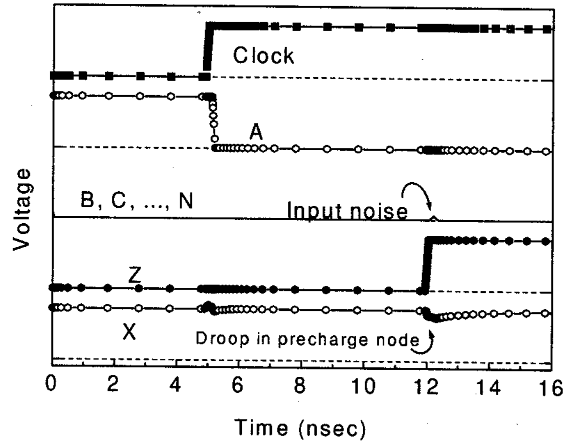

Fig. 16 **The input and output waveforms for the circuit shown in Figure 15.**

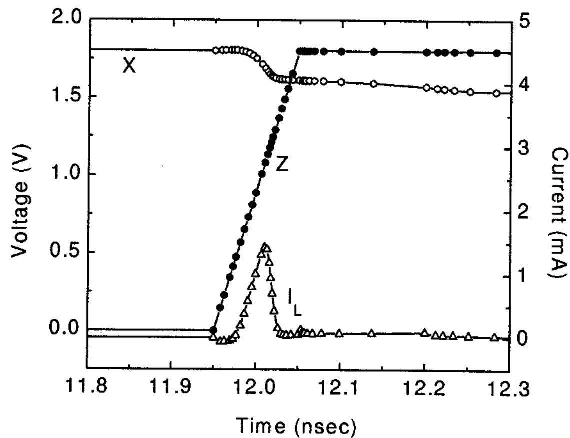

Fig. 17 **Same as Figure 16, except the area of detail is enlarged, and the transient leakage current is shown.**

Yet another solution is to re-order the NFET logic tree. Here, the largest parallel group of NFETs is placed at the bottom of the stack, therefore preventing their bodies from charging to high voltage. Again, this solution may lead to some degree of charge sharing problem.

Finally, a conditional feedback circuit can be provided. This circuit will help the PFET keeper device at severe conditions (such as burn-in elevated voltages) hold the precharge node at its high value. The feedback circuit will be disconnected during regular use conditions, so its impact on circuit delay will be minimal.

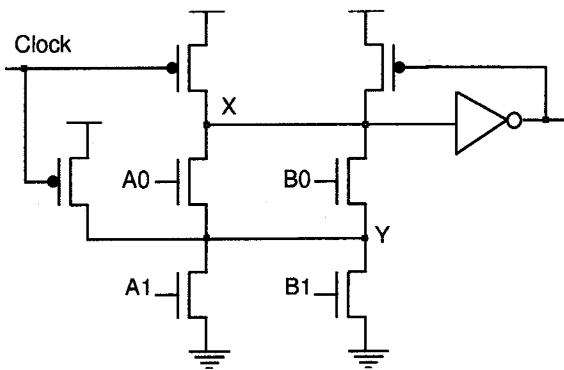

Fig. 18 **Original bulk design using pre-charging of the intermediate node Y.**

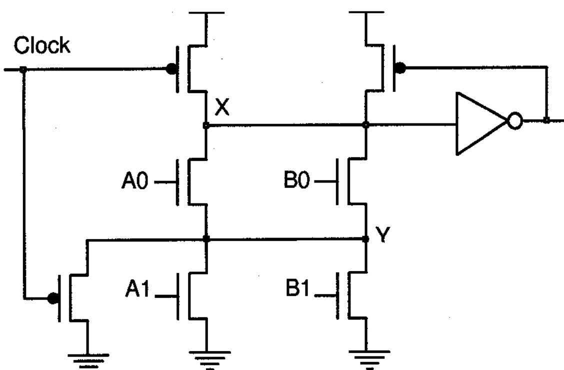

Fig. 19 **SOI modification of Figure 18. The intermediate node Y is pre-discharged.**

SRAM ISSUES

Above discussions focused on the impact of pass-gate transient leakage current on noise margin of dynamic circuits. However, this leakage can impact other circuits as well. For example, it can become problematic for SRAM arrays during the read and write operations. This SRAM failure has both spatial and temporal dependence. In other words, it is sensitive to both bit patterns and to timings: Imagine that all the 512 cells on the bitline have stored 1's. Now, if the wordline for the cell 0 is selected and we try to write a zero to this cell, all the other 511 cells will develop the pass-gate leakage current and will fight the write operation.

To address this issue and a similar one that arises during the read operations, several circuit solutions can be used: The bitline can be charged to voltages less than V_{dd}, such as V_{dd}-V_t. The size of the write switch and the driver can be increased. One can limit the number of cells on the bitline. Finally, if the fail is limited to high voltages used for screening or testing, the test/screen pattern can be modified.

So far, all of the discussions have been limited to use of floating body devices only. We have not described the use of body-contacted devices as solutions to previous problems. Body-contacted devices are much slower than floating-body devices, they take more space, and they add a large amount of extra capacitance. Therefore, their implementation in SOI should be truly a last resort solution. One case that may warrant this solution is the sense amplifier design for the SRAM arrays. Here the V_t-matching of sense devices plays a crucial role in functionality and speed of the amplifier. Therefore, no V_t variation due to history effect can be tolerated and body ties should be used.

3. CONCLUSION

In summary, we described specific circuit issues that may arise in design of CMOS VLSI on PD SOI. We analyzed the problems and explained some of the possible solutions. Since SOI circuit advantages over bulk have been described well elsewhere, we did not specifically address them here but will review them at the tutorial.

4. ACKNOWLEDGEMENT

We gratefully acknowledge detailed simulation work by K. Wu, and many useful discussions with G. Shahidi, D. Cox, S. Storino, L. Wagner, M. Michail, M. Ketchen, M. Shernoy, J. Sleight, J. Warnock, S. Fung, S. Lo, R. Schulz, W. Rausch, C. T. Chuang, G. Sai-Halasz, G. Smith and many other IBM colleagues.

5. REFERENCES

[1] G. G. Shahidi, et. al., "SOI for a 1-Volt CMOS Technology and Applications to a 512-Kb SRAM with 3.5nsec Access Time," in *Tech. Dig. IEDM,* 1993, pp. 813-816.

[2] G. G. Shahidi, et. al., "A Room Temperature 0.1μm CMOS on SOI," *IEEE Trans. Elec. Dev.*, vol. 41, no. 12, Dec. 1994, pp. 2405-2412.

[3] F. Assaderaghi, et. al., "A 7.9/5.5 psec Room/Low Temperature SOI CMOS," in *Tech. Dig. IEDM*, 1997, pp. 415-418.

[4] D. Suh and J. G. Fussom, "Dynamic Floating-Body Instabilities in Partially Depleted SOI CMOS Circuits," in *Tech. Dig. IEDM*, 1994, pp. 661-664.

[5] D. J. Schepis, et. al., "A 0.25 μm CMOS SOI Technology and Its Application to 4Mb SRAM," in *Tech. Dig. IEDM*, 1997, pp. 587-590.

[6] K. Mistry, et. al., "A 2.0V, 0.35μm Partially Depleted SOI-CMOS Technology," in *Tech. Dig. IEDM*, 1997, pp. 583-586.

[7] D. H. Allen, et. al., "A 550MHz 64b SOI Processor with Cu Interconnects," in the 1999 *ISSCC.*

[8] M. Canada, et. al., "A 580MHz RISC Microprocessor in a 0.12μm Leff CMOS SOI Technology," in the 1999 *ISSCC.*

[9] J. P. Eckhardt and P. D. Meunch, "A SOI Specific PLL for 1GHz Microprocessors in 0.25μm 1.8V CMOS," in the 1999 ISSCC.

[10] Y. Chang, et. al., *IEEE TED*, pp. 277-287, 1997.

[11] S. Voldman, et. al., "Dynamic threshold body- and gate-coupled SOI ESD protection networks," *Journal of Electrostatics*, Sep. 1998, pp. 239-257.

[12] F. Assaderaghi, et. al., "Accurate Measurement of Pass-Transistor Leakage Current in SOI MOSFETS," *IEEE SOI Conf.*, 1996, pp. 66-67.