Device and Circuit Design Issues in SOI Technology
(Invited)
Ghavam G. Shahidi, Atul Ajmera, Fariborz Assaderaghi, Ronald J. Bolam,
Harold Hovel, Effendi Leobandung, Werner Rausch, Devendra Sadana,
Dominic Schepis, Lawrence F. Wagner, Larry Wissel, Kun Wu, and Bijan Davari
IBM Semiconductor Research and Development Center, Hopewell Junction, NY 12533, USA
Phone: (914)-892-2075, Fax: (914)-892-2568, shahidi@us.ibm.com

I- Abstract

Partially-Depleted deep sub-micron CMOS on SOI technology is becoming a mainstream technology. This technology offers 20-35% performance gain over a bulk technology implemented with the same lithography. This paper first reviews the partially-depleted SOI device and describes reasons why it was chosen over fully depleted SOI device. Next the sources of performance gain on SOI are reviewed. SOI-unique circuit and technology issues that a designer must consider and account for are discussed next. Finally, a low-power application of SOI is reviewed.

II- Introduction

Deep sub-micron room-temperature bulk CMOS is the main technology for ULSI systems. CMOS scaling, which has been the main tool for improving system performance, is rapidly approaching its limits. Use of SOI for improving CMOS performance has been under consideration for nearly three decades [1]. The main challenges in acceptance of CMOS on SOI have been: material quality; device design on SOI; bulk scaling and the resulting performance gain per generation; and the lack of demonstration of a mainstream complex application implemented in SOI.

Over the last few years, significant progress has been made in all of the above areas. The first major breakthrough was adoption of the partially-depleted (PD) SOI device [2]. As we will discuss, PD SOI significantly simplifies the device and circuit design on SOI (as compared to a “fully-depleted” SOI device). In fact, the “partially-depleted” condition is one major source of SOI performance gain over bulk CMOS. The choice of a PD device introduces a few new circuit behaviors that do not exist in bulk technology. The key effects are the “kink effect”, “history dependence” of the delay, “pass-gate leakage”, and “self-heating”. The effects need to be accurately measured and modeled; their effects in circuits must be well understood; and, when needed, design techniques must be modified to account for these effects.

SOI technology is also being considered for low-power applications. In the last section, aspects of the SOI technology which make it attractive for low-power are discussed.

III- Device Design on SOI

Figure 1 shows the cross section of an NMOS transistor on SOI and its equivalent circuit. The main feature of MOS on SOI is that the device’s local substrate (“body”) floating, and its bias depends on the “history” of the device. The changing source-to-body bias V_{BS} in turn modulates V_{T}. One method proposed to minimize the floating body effects has been the use of fully-depleted (FD) SOI. In FD devices, the SOI

Figure 1: NMOS on SOI and its equivalent circuit

15.1.1
film thickness is much smaller than the depletion width of the device, and its potential is tightly controlled by the gate. In other words, there is no neutral region of the body of the MOS that can be charged. The “kink effect” is a phenomenon where the output conductance increases as impact ionization starts and $V_{GS}$ becomes positive for $V_{DS} > 1.1$ V). The “kink effect”, which is a major cause of the performance boost of digital circuits on SOI because of the increase in current, is present in PD SOI.

Another perceived benefit of FD SOI has been improvement in short-channel effects (SCE) [1]. Figure 2 is a simulation of $V_T$ roll-off as the SOI film thickness is reduced (bottom four curves), along with the roll-off on bulk (top curve). As the SOI film thickness is reduced, the $V_T$ roll-off improves. The improvement in roll-off is caused by the reduction in junction depth. As shown in Fig. 3, if the junction depth on bulk is the same as the SOI film thickness, then the roll-off on bulk and SOI is the same. Indeed, FD SOI does not result in an improved SCE. By going to a PD device, one can use many of the techniques developed in bulk CMOS for controlling the SCE (i.e. retrograde well, extension, and HALO).

<table>
<thead>
<tr>
<th>Manufacturability</th>
<th>PD</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design Point (high $V_T$)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Multiple $V_T$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SCE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Kink Effect</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Body Contact</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Pass Gate Leakage</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>History Dependence</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1 - PD vs. FD SOI Device

There are many other convincing benefits associated with PD SOI, as summarized in Table 1. PD SOI is much easier to manufacture, since the SOI film is thicker (150 nm compared to a FD thickness much less than 100 nm). A high $V_T$ can be achieved on PD SOI, whereas it is very difficult to obtain a high $V_T$ on FD SOI: if the film doping is increased to increase the $V_T$, the device becomes PD; if the film thickness is reduced to make it FD, then $V_T$ is reduced. Deep-submicron CMOS technologies use multiple $V_T$ values to balance performance against noise and power. It is possible to have a multiple-$V_T$ offering in a PD SOI technology, but not on FD SOI. The kink

Figure 4: Cross-section of SOI CMOS with Cu metallization
effect is present on both PD and FD SOI. But as will be shown later, the kink effect (the fact that \( V_{DS} \) is positive most of the time) is a major source of the performance advantage of bulk over SOI. There are cases where the floating-body effects must be eliminated at all cost; in those cases, it is possible to form a body contact on SOI, but a body contact is not feasible on FD SOI. "Pass gate" leakage is higher on FD devices (since the bipolar gain of the device is higher). "History dependence" of the delay is larger on PD. But, as it will be shown later, this is a manageable phenomenon in most circuits. In summary, there are very convincing reasons why the PD device design is preferred.

IV- Device Results

Floating-body effects, which are proportional to supply voltage, are reduced at lower voltages. As the mainstream bulk supply voltage is dropped below 1.8 V, a complete CMOS on SOI technology becomes feasible (including a high-voltage defect-screening methodology as part of test). Initially, a 0.18 \( \mu \)m CMOS (gate lithography) on SOI with aluminum metallization was developed [3; "0.25 \( \mu \)m" in the title refers to gate lithography], followed by a 0.22 \( \mu \)m CMOS SOI with copper metallization. Figure 4 is the cross section of the 0.22 \( \mu \)m CMOS with six layers of copper.

Fig. 5 is the \( I_D-V_{DS} \) characteristics of both NMOS and PMOS from 0.25 \( \mu \)m SOI CMOS. The kink effect results in lower output resistance. The buried oxide layer in SOI is a good electrical insulator. Thus if a device is continuously "on" (as during I-V measurements), its temperature will go increase, resulting in reduced current (solid curves in Fig. 5). The dashed curves in Fig. 5 are the I-V characteristics without self heating [4]. The curves "without self-heating" more closely represent the actual switching current in most circuits.

Fig. 6 is the subthreshold current of the device. The SOI MOSFETs have larger drain-induced barrier lowering (DIBL) and lower saturation \( V_T \) (at nominal channel lengths), caused by the floating body effect.

Circuits in SOI operate faster than identical circuits in a bulk technology of the same lithography. There are three primary causes of performance gain of SOI over bulk: 1. Area junction capacitance is nearly absent on SOI. 2. The classic "MOS reverse body effect" is absent on SOI, because the body-source bias can not be negative (as is the case in stacked transistors and pass gates in bulk). Indeed in SOI,
$V_{BS} > 0$ (which gives rise to floating body effects).

3. Because of positive $V_{BS}$ in most switching cases, SOI “on” current is higher than bulk. Figure 7 is the ‘on’ current of nFET and pFET as a function of $L_m$, with and without self-heating. The bulk CMOS currents are close to the SOI currents with self-heating. SOI device “on” without self-heating current is 10-15% higher than bulk CMOS.

Unloaded inverters on SOI are 20-25% faster than their bulk counterparts (more “on” current; less junction capacitance). For NAND2, NAND3, and NAND4, the performance gain on SOI goes from 27% to 50% because the floating body in SOI does not cause $V_t$ reduction as the number of series NMOS transistors increases. Table 2 lists the simulation of critical paths of a number of microprocessors, using the models for 0.22 μm CMOS-SOI. References 5 and 6 describe microprocessors built using this SOI technology.

<table>
<thead>
<tr>
<th>Circuit</th>
<th>($t_{BULK}-t_{SOI}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU1 L2 Directory</td>
<td>29.8%</td>
</tr>
<tr>
<td>Access time</td>
<td></td>
</tr>
<tr>
<td>CPU2 (Gate dominated)</td>
<td>27.1%</td>
</tr>
<tr>
<td>CPU2 (Wiredominated)</td>
<td>20.9%</td>
</tr>
<tr>
<td>CPU3 (Gate dominated)</td>
<td>30.4%</td>
</tr>
<tr>
<td>CPU3 (Wire dominated)</td>
<td>19.6%</td>
</tr>
</tbody>
</table>

Table 2: Simulation of critical paths of CPUs

V- Floating Body Effects

One of the challenges of PD SOI CMOS technology is understanding, characterization, modeling and circuit design in the presence of floating body effects. Figure 8 shows the circuit elements determining the $V_{BS}$ of a SOI FET. The body voltage is determined by the p-n diode leakages and the impact ionization current (Fig. 8(a)) and capacitive coupling to external nodes during switching (Fig. 8(b)). Response time of the leakage currents is slow (of the order of ms) and the response time of capacitive coupling is fast (ps). The change in $V_{BS}$ modulates $V_t$ according to Fig. 9. In PD SOI, $V_{BS} > 0$ in most cases. This is in contrast to bulk where $V_{BS} < 0$ all the time. If it is desired to avoid the floating body effects at all cost, one can use direct contact to the MOS local substrate, either grounding it, or shorting it to the source ($V_{BS} = 0$) [7]. Figure 10 shows the $I_{C-VDSS}$ curves of an nFET with body left floating (dashed curves, showing the “kink effect”) and body grounded (solid curves). When the body is shorted to the source, the output resistance improves significantly, which is desirable for analog circuits.
VI- History Dependence

There are two SOI-unique effects to consider when designing in SOI or migrating a design from bulk to SOI. "History dependence" is the change in delay through a gate as a function of the switching history of the device [8]. This effect is illustrated in Fig. 11. As the period of input pulse changes (Fig. 11(a)), the delay through the gate changes. Fig. 11(b) is the delay through an even number of NANDs. During fast switching, the $V_{BS}$ is determined by its capacitive coupling to the gate, drain and source voltages. The capacitive coupling is on top of the "steady-state" value of the $V_{BS}$. In "steady state", $V_{BS}$ is determined by the diode leakage and the impact ionization, as in Fig. 8(a). It is this variation of "steady state" value of the $V_{BS}$ that gives rise to variation in $V_T$ and the delay. This is illustrated in Fig. 12. When the gate of a nFET in an inverter turns "on", initially the body capacitively couples to the drain and is pulled low. Then slowly it converges to a value set by the drain, source and gate voltages as well as leakage. If, during the "slow" return of the body potential to its initial value, another switching event happens, then the delay will vary. Figure 13 shows the device "on" current measured by pulsed I-V. The difference in the I-V curves when the period is 1 $\mu$s or 1 ms is caused by the history effect.

Indeed the variation in delay through a gate can be measured as the input period is changed. Figure 14 shows the variation in the delay through the gate as the pulse period is changed. The shape of the curve agrees fairly well with the variation in body voltage. When the body bias when the gate is turned on (as in Fig. 12) is plotted against time on log-axis, the behavior in Fig. 15 is similar to that of Fig. 14. For short periods of time (i.e., less than 1 $\mu$s), $V_{BS}$ is fairly

![Figure 13: Variation of ON current caused by the device on cycle time](image)

![Figure 14: History dependence of the delay](image)
stable. But for intermediate times (i.e., between 1 \mu s and 1 mS) V_{SS} varies, and for longer times, V_{SS} reaches its steady state.

Table 3 lists the maximum variation in delay observed for a number of gates due to variation in the input signal period:

![Figure 15: Variation of V_{SS} when the device turns ON](image)

<table>
<thead>
<tr>
<th>Circuit</th>
<th>% Change in Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverter</td>
<td>5%</td>
</tr>
<tr>
<td>NAND2, Bot. Device</td>
<td>3%</td>
</tr>
<tr>
<td>NAND2, Top Device</td>
<td>5%</td>
</tr>
<tr>
<td>NAND4, Bot. Device</td>
<td>4%</td>
</tr>
<tr>
<td>NAND4, top Device</td>
<td>7%</td>
</tr>
<tr>
<td>Inv. + 2mm Wire</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 3: Variations in Delay

The uncertainty in the delays in Table 3 may seem excessive. However, there is considerable uncertainty in the delay through a gate in a bulk design also. Among the phenomena contributing to uncertainty in a bulk design are: intra-chip process variation (about 15-20% of a gate delay); delay variation caused by top vs. bottom switching device in NAND gate, or the number of simultaneous “1”s arriving at a NOR gate (about 20-35%); on-chip V_{DD} variations (about 10%); and temperature variations (on-chip and ambient, which can cause 10-20% change in delay). The SOI-induced uncertainty in delay, while not negligible, is no larger than the other sources of uncertainty in a chip, and can be handled in a similar fashion [9].

VII- Pass-Gate Leakage

Another key SOI unique effect is pass-gate leakage [10]. In SOI if source and drain are both high, the body can charge up all the way to V_{DD}, and if the gate is “off”, then the body will accumulate after a long interval (i.e., ms). Now if the source is pulled low (Fig. 16(a)), the holes (in nFET) will go to the source, and inject electrons to the drain. This will give rise to a pulse of current in the device, even though the gate is “off” (Fig. 16(b)). This effect can lead to the discharge of nodes with no keeper devices (such as in DRAM cells), or with a very weak keeper device (such as in a wide dynamic OR). The case of wide OR is illustrated in Fig. 17. If V_{DD} across all of the devices A_i is high for a long time, then they will accumulate. Now if device C is turned on (i.e., the common source of A_i, devices is pulled low), then a pulse of current will accumulate.
be passed through the devices A. If the A devices are wide enough, they can discharge the dynamic node (as shown in Fig. 18). Possible solutions to this problem are: increasing the size of the keeper device (which results in loss of performance); limiting the logical width of the OR function; eliminating the “charge-sharing” pFET; or discharging the common source of A, devices periodically [5,10].

Figure 18: Pass-gate leakage induced fail [4]

VIII- “Self-Heating”

One of the concerns about SOI is the presence of the buried oxide and the subsequent reduction in thermal conductivity to the substrate. The decrease in thermal conductance results in an increase in device temperature, with the subsequent reduction in current and possible impact on reliability. The thermal resistance to the substrate is 60-100 μm C/mW. One way to estimate the temperature rise in a device is to calculate the power dissipation in a given device per unit width. In a 0.22 μm CMOS at 1.8V, a continuously “on” NFET has a current near 600 μA/μm, dissipating about 1 mW and resulting in a 60-100 °C temperature rise. In a switching circuit, the device conducts current only during during falling (or rising) transients. Furthermore, the device does not see the full VDD throughout of the transient. The low duty cycle of power dissipation is illustrated in Fig. 19. The power per unit width of the device, even in the clock drivers, is much less than 1 mW/μm. The temperature rise in most devices is proportionally small, about 2-3 °C. In terminated output drivers, where the device actually drives a resistive load and conducts current for a considerable fraction of the cycle, the temperature increase is about 10-15 °C.

Figure 19: Self-heating in nFET

IX-SOI for Low Power

One of the attractions of the SOI technology is its low-power behavior. It has two origins: for a given CMOS generation, SOI has higher performance than a comparable bulk technology. This performance “headroom” allows for operation at lower voltage, and much-lower power. This effect is illustrated in Fig. 20, where power is plotted against delay for an unloaded NAND3 delay chain for a number of technologies (0.30 μm to 0.22 μm bulk CMOS, and 0.22 μm SOI CMOS). As the performance of the CMOS technology improves, its power is reduced.
The other attraction of SOI is that $V_{BS} > 0$ in most switching cases. Positive $V_{BS}$ dynamically lowers $V_T$, and a lower $V_T$ is valuable at low voltage operation where the performance and functionality is proportional to $V_{DD}$ minus $V_T$. Figure 21 is power vs. delay for a 4 Mb SRAM at different $V_{DD}$ values for both bulk and SOI. At a given delay, the SOI implementation has two to three times less power than the bulk implementation.

**X-Summary**

We have described a PD SOI CMOS technology, with considerable advantage over FD technology in terms of design point and manufacturability. We described the sources of SOI performance advantage, enabling SOI CMOS to achieve a 20-35% performance advantage over bulk CMOS. The floating-body CMOS topology introduces a number of unique effects that can impact circuit functionality. They are kink effect, history dependence of the delay and the pass gate leakage. All these effects are manageable, and circuits can be designed around them. We also discussed the attributes of SOI valuable to low-power applications.

**XI-References**