This paper describes applications of Silicon on Insulator (SOI) technology to high performance on-chip memories (e.g. SRAMs and register files etc.). The primary focuses are on the important and unique issues in SOI technology such as performance gain, history effect, power reduction, pulsewidth control and self-heating. The effects on the interconnect performance and reliability are also discussed. The hardware data show that Partially Depleted (PD) SOI technology results in 20-25% and 18-22% improvement in the performance over bulk cmos using 0.25μm and 0.18μm technology nodes respectively. Power reduction of 7% to 8% over bulk technology is realized. Self-heating in SOI can cause device and interconnect temperatures to rise more than 100-120°C.

I. INTRODUCTION

As the scaling in bulk silicon technology approaches its limits, SOI has emerged as a technology for mainstream digital applications to improve performance further (i.e. to reduce "RC" further)[1-3]. Due to reduced junction capacitances and "dynamic threshold voltages" SOI renders higher performance than bulk. Partially depleted SOI (Fig. 1) offers smaller source-to-drain series resistance and dynamic threshold voltages which result in higher performance over fully depleted SOI technology.

Also the process and device designs for partially depleted SOI are much more compatible with the bulk CMOS. However, the floating body in partially depleted device can accumulate or lose charges. This may cause high device leakages during standby or the high temperature burn-in conditions. The amount of charge in the body depends on the conditions of source, drain and gate of the device. This results in variable body potentials and dynamic threshold voltages which cause
variable device delays known as "history effect". One of the factors which affects the history effect is the interconnect loading. Also poor thermal conductivity of buried oxide in SOI can lead to self-heating which can impact devices and interconnects. If organic low-k dielectrics are explored to improve the performance further, these dielectrics may suffer from self-heating due to poor thermal conductivity and the effect of self-heating on them is unknown. This paper highlights performance and power advantages of SOI over bulk technology through specific macro elements used in microprocessor (e.g. Register file, SRAM, clock buffers etc) as well as the interrelation between SOI and interconnect.

**II. BULK and SOI PERFORMANCE**

Recently a high-speed L1 cache SRAM directory (34 Kb) with read access time below 430 ps and a frequency of 2 GHz in 1.5V, 0.18μm CMOS bulk technology has been demonstrated [4]. The key elements for high performance are new circuit techniques; front end and Cu back end technology. The cross-section of the SRAM with full 7 levels of Copper interconnect is shown in Fig. 2.

![Cross-section of SRAM with Cu interconnects](image)

**Figure 2: L1 directory SRAM with multilevel of Cu interconnects.**

The array macro uses 3 levels of metal M1-M3 plus tungsten as local interconnect. Array bitlines are on M2, wordlines are on M3; while levels M4-M7 are used for unit and chip level global wiring, power distribution and I/O connections. The local clock skews are minimized using optimum wire lengths and widths to assure proper timing. The reduced heights of Cu wire with lower resistance is beneficial in reducing bitline interconnect loads. Such SRAMs with proper circuit tuning can be mapped into SOI with improvements close to 18-22% due to reduced junction capacitance and dynamic threshold voltages. An aggressive 6-T SRAM cell in SOI including interconnect features is shown in Fig. 3. The derivatives of such cells are used in high-speed memory units, which are part of the data paths (e.g. register files). Here performance improvement for SOI is shown over bulk. The read access time is plotted as a function of temperature in Fig. 4. The improvement in the access time is around 3.4% per 10°C temperature drop compared to 1.6% for per 10°C temperature drop for the bulk. This performance improvement can be attributed
quantitatively to two factors - junction capacitance reduction and dynamic threshold voltage (Vt) changes. In case of SOI total capacitance consists of wire and gate capacitance while in case of bulk it comprises of junction capacitance as well as wire and gate capacitance.

By comparing the slopes of power-frequency curves for bulk and SOI chips at a fixed supply voltage (V_{dd}) the reduction in the device capacitance can be obtained. This value is directly related to the performance gain. From the slopes a reduction of 7-8% in junction capacitance is observed for SOI chips (Fig. 5). The rest of the gain is attributed to dynamic Vt of SOI due to charge accumulation. Out of the total power, interconnect power accounts for about 10%. Using the same analysis it is noted that the reduction in capacitance (i.e. derived from the slopes of power frequency curves changes marginally from room temperature down to sub-zero temperatures. This indicates that the reduction of junction capacitance in the case of SOI remains unchanged across the temperature. Similar reduction in capacitance is observed for a full CPU chip. In the case of SOI most of the capacitance is due to gate and wire, and thus SOI performance gain can be smaller if the critical paths are heavily wire dominated.

III. IMPACT OF INTERCONNECTS ON HISTORY EFFECT

History effect is another characteristic of SOI technology. History effect depends on many variables such as supply voltage, output loading, slew rate, duty cycle, input pulsewidth and temperature. The most important factors related to interconnects are output loading and slew rates.
If the macro sees large wire loads ("RC") or device loads then history effect can be significant. This is shown through access time vs frequency graph as a function of output load (Fig. 6).

![Figure 5: Power vs. frequency for bulk and SOI for multi-port register file.](image1)

![Figure 6: History effect as a function of interconnects loading for multi-port register file.](image2)

The variation in access time for 0.13 pf loading is 6.76% (defined as a history effect), while for the larger output load of 0.43 pf it goes up to 7.2%. At large loads the signal rise and fall times are slowed down, thus increasing the more cross-over region during switching of nfets and pfets and resulting in more impact ionization current (and hence more charges injected into the body). On the other hand there is less time to replenish charge at higher frequencies, so the variations in access time are more pronounced at higher frequencies. For similar reasons the output pulsewidth variation increases from 5.6% to 6.7% due to increased loading (not shown). Such effects are taken into account for the timing of the macro.
IV. SELF-HEATING EFFECT ON INTERCONNECT PERFORMANCE

Another important aspect of SOI is self-heating and its impact on interconnects and devices. To illustrate this the steady state temperature distribution of a multi-finger NFET with a realistic physical geometry in a clock buffer was studied using a commercial 3-D finite-difference thermal code [5]. The geometry of a four-finger FET on a chip including the body, source and drain MCBAR as a local interconnect, gate polysilicon, five layers of Copper metal, vias and dielectric was modeled in cartesian coordinates. Junction and level 1 metal (M1) peak temperature rise in the SOI device with one active finger is shown in Figure 7. Only the leftmost channel is active and shows the highest temperature rise. The normalized junction-to-chip temperature rise or thermal resistance is $50^\circ$C-$\mu$m/mW ($3^\circ$C rise for 0.06 mW/$\mu$m heat dissipation). Values reported in the literature range from 50-90$^\circ$C-$\mu$m/mW for single-finger devices with similar dimensions [6,7]. Figure 8 illustrates the temperature distribution in all the metal interconnect (MCBAR to M5) with four fingers active. With multiple fingers active, the junction temperature rise increases as shown in Figure 9, even though the heat dissipation per unit channel width remains the same (0.06 mW/$\mu$m). This is due to the close thermal coupling between fingers within the same body.

![Figure 7: Junction and level 1 metal (M1) peak temperature rise in a four-finger FET with only the leftmost finger active.](image)

Our results clearly show that the junction-to-chip thermal resistance increases by more than a factor of three as more fingers become active. These results show that the conventional use of a single thermal resistance in SPICE type models for a SOI device without differentiating between a single-finger or a multi-finger structure is erroneous. In particular, the use of a thermal resistance value derived or measured from a single-finger isolated device would fail to properly account for the temperature rise of multi-finger devices due to self-heating. Furthermore, SPICE models can not account for both types of geometries simultaneously for thermal calculations. By comparison, the temperature rise in a similar bulk device is small (Fig. 9) and the circuit simulation errors are negligible. Circuit performance was evaluated using the clock tree from a real microprocessor shown in Figure 10. Performance changes with respect to the bulk CMOS performance at 25$^\circ$C of SOI and bulk devices using a dielectric constant of 4.0 are shown in Fig. 11. For the chosen $L_{eff}$ and $V_{dd}$ (1.8 V) the performance gain is over 20% for SOI in the temperature range of 30 - 80$^\circ$C and reduces at lower temperatures. The performance change for SOI and bulk per 10$^\circ$C change in
junction temperature is 1.2% and 1.32% respectively. Also the increased temperature adds to the degradation in reliability of SOI and interconnects (not shown here).

Figure 8: 3-D thermal analysis of metal interconnects for a four-finger FET with all the four fingers active.

Figure 9: Multiple fingers vs temperature rise.

Figure 10: Clock buffer in a microprocessor.

Figure 11: Clock buffer performance vs. temperature.
V. SUMMARY

Recent advances in high speed L1-cache directory SRAMs with Cu interconnects are described and applications of SOI technology to high speed memories are demonstrated. The impact of SOI on interconnect performance and reliability are discussed. With the inherent advantages of power reduction and improvement in performance, SOI remains as a viable device technology for high performance processors. By optimizing the interconnect and device loadings, the history effect can be minimized.

REFERENCES


4. R. V. Joshi et al., "A 2 GHz cycle, 430 ps access time 34 Kb L1 Directory SRAM in 1.5 V, 0.18µm CMOS bulk technology", Symp. VLSI Ckts., June 2000, pp. 222-225.

