Circuit-aware Device Design Methodology for Nanometer Technologies: A Case Study for Low Power SRAM Design

Qikai Chen, Saibal Mukhopadhyay, Aditya Bansal and Kaushik Roy
School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA
{qikaichen, sm, bansal, kaushik}@purdue.edu

Abstract:

In this paper, we propose a general *Circuit-aware Device Design* methodology, which can improve the overall circuit design by taking advantages of the individual *circuit characters during the device design phase*. The proposed methodology analytically derives the optimal device in terms of the pre-specified circuit quality factor. We applied the proposed methodology to SRAM design and achieved significant reduction in standby leakage and access time (11% and 7%, respectively, for conventional 6T-SRAM). Also, we observed that the optimal devices selected depend considerably on the applied circuit techniques. We believe that the proposed *Circuit-aware Device Design* methodology will be useful in the sub-90nm technology, where different leakage components (subthreshold, gate, and junction tunneling) are comparable in magnitude. Also, in this work, we have presented a design automation framework for SRAM, which is conventionally custom designed and optimized.

1. Introduction

Technology scaling results in significant leakage current increase in CMOS devices. In nano-scale CMOS devices, the major components of leakage current are the sub-threshold leakage, the gate direct tunneling leakage and the reverse biased band-to-band-tunneling junction leakage (Fig. 1). These different leakage components are strong functions of device geometry and doping profile. Also, in nano-scale devices, these leakage components are comparable in magnitude [1]. Therefore, the device design methods used to optimize one particular leakage component may increase another one.

SRAM caches dominate the chip area of the state-of-the-art microprocessors (predicted to occupy about 94% of die-area by 2014 [2]). Since a large portion of SRAM cells in the cache always remain in the stand-by mode, leakage power dominates the total power consumption. Furthermore, the gap between the memory access time and the processor speed also increases with the technology scaling. Hence, designing SRAM cells with low access time and low leakage is of great importance in nano-scaled technologies. Several circuit techniques, such as, source-biased SRAM (SBSRAM), body-bias SRAM (BB-SRAM), etc., have been proposed for low leakage (Fig. 2) [3]. Such low leakage circuit techniques primarily reduce the sub-threshold leakage (SBSRAM, BB-SRAM) and to some extent the gate leakage (SBSRAM). However, they may adversely impact other leakage components. For example, a reverse-body-bias in the SRAM stand-by mode (RBB-SRAM) actually increases the junction tunneling band-to-band leakage. Since, in nano-scale devices, power consumptions due to each of the three leakage components are comparable [4], the effectiveness of the different circuit technique depends considerably on the relative magnitudes of each leakage component in a device.

Conventionally the design solutions have been sought *independently* from the device and the circuit levels. In the device level, geometry and doping profile of the transistors are optimized to maximize the "ON" current (Ion) while minimizing the "OFF" (Ioff) current (i.e. to maximize the Ion/Ioff ratio). In the circuit level, different low power circuit solutions are optimized through transistor sizing, proper selection of source-bias or body-bias voltages as in SBSRAM or BB-SRAM, and etc. Then, low power circuit technique is applied to SRAM design with the separately optimized device. Fig. 3a illustrates the different steps of the above mentioned design flow (hereafter, referred to as Individually Optimized Device-Circuit Design (*IndOptDevCirc-design*)). It should be noted that, the design flow shown in Fig. 3a is not particular to SRAM and can also be generalized for logic circuits.

Although, the *IndOptDevCirc-design* flow improves the quality of a circuit design, it does not efficiently take advantage of the impact of different low-leakage circuit techniques on different leakage components. In technologies larger than 100nm, where sub-threshold leakage is the only dominant leakage component, the circuits designed with transistors having minimum device leakage also results in the minimum overall circuit leakage. Hence, the *IndOptDevCirc-design* flow is effective in designing low power circuits for such technologies. However, in nano-scaled devices, where, all the leakage components mentioned earlier are comparable in magnitude, the consideration of the circuit properties provides more opportunities for design optimization. Thus, a design effort that simultaneously considers circuits and devices options is expected to be more effective in achieving low-leakage while maintaining high-performance.

In this paper, we propose a device design methodology, Fig. 3b, which efficiently takes advantage of the circuit techniques by simultaneously considering device and circuit issues (hereafter

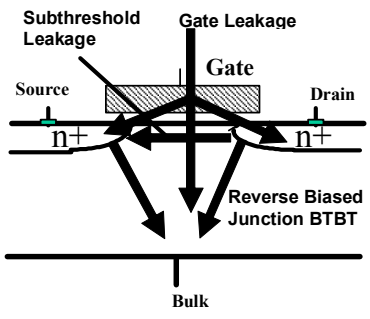
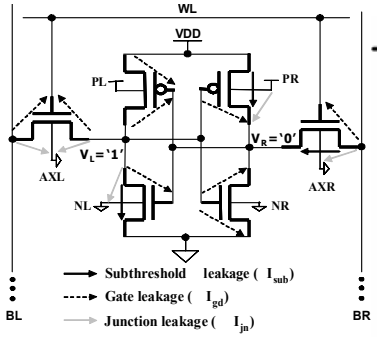
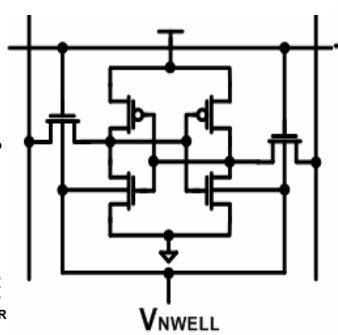


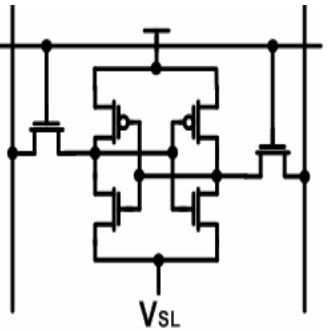
Fig. 1: Leakage components in a scaled transistor



(a) Conventional 6T-SRAM



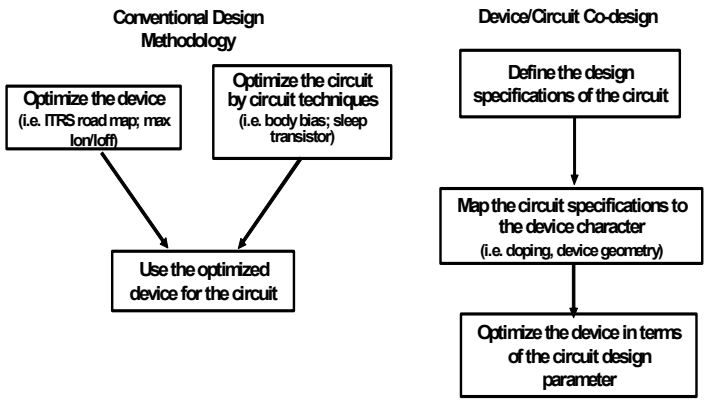
(b) BB-SRAM



(c) SB-SRAM

Fig. 2: SRAM Circuit

3-9810801-0-6/DATE06 © 2006 EDAA



(a) IndOptDevCirc-design **(b) Circuit-aware Device Design**

Fig. 3: Different device design flow

referred to as *Circuit-aware Device Design*). In this flow, Fig. 3b, first the circuit quality factor, which should be maximized for the optimal design, is defined in terms of the weighted performance, power consumption and etc. Then, circuit characters (i.e. performance and power) are represented by device properties (such as ON current, leakage current, junction capacitance, and etc.). Finally device properties are transformed down to the device design parameters, such as doping and geometry. By doing so, based on the quality factor from step one, the device optimization goal regarding the particular circuit is clearly defined. The optimal device can be derived analytically following the proposed flow. It should be noted that, the proposed flow allows co-optimization of the devices and circuits. The circuit parameters, such as sizing, involved in defining the quality factor can also be optimized along with the devices to further improve the design quality.

In the following sections, we have used the SRAM design at 65nm technology node [2] as a case study to show the advantage of *Circuit-aware Device Design methodology*. Section 2, 3, and 4 correspond to the 3 steps in the proposed design flow as shown in Fig. 3b. It should be noted that the devices can be optimized through doping profile and device geometry. However, to simplify the problem, in this paper we considered only the doping profile optimization. The main contributions of this paper are:

- We propose a *Circuit-aware Device Design* methodology, which analytically selects the right device design with the information of design specifications and circuit styles.
- We develop an automation design flow for SRAM, which is conventionally custom designed. With the proposed SRAM design automation framework, the optimal device can be selected analytically. This reduces the design time.
- We apply the proposed methodology to SRAM design in 65nm technology based on ITRS road map [2]. We optimize the doping profile of devices so as to improve the power and performance of a conventional 6-T SRAM, a SB-SRAM and a BB-SRAM.

The principal observations from the *Circuit-aware Device Design* methodology are:

- The device that results in optimum leakage and performance in a conventional 6-T SRAM cell is different from the device optimized for maximum Ion/Ioff ratio. The *Circuit-aware Device Design flow* results in 11% lower leakage and 7% higher performance for a conventional 6-T SRAM.
- The doping profile that results in optimum leakage and performance in conventional 6-T SRAMs, SB-SRAMs and BB-SRAMs are significantly different from each other.

From our observation, we believe that the proposed design flow and the presented analysis will be useful for designing low-power, high-performance circuits in nanometer regime, especially SRAM designs as demonstrated in this paper.

2. Modeling of SRAM Design Specifications

In the proposed design methodology, the first step is to define the circuit optimization goal, referred to as quality factor in this paper. In this section, each parameter in the quality factor is mapped to the circuit level properties, such as capacitance and current values.

In SRAM (Fig. 2a), there are two important design specifications -- power consumption and access time. Due to the large size and low activity, the power consumption of SRAM is dominated by the leakage power of the individual cells. Similarly, the access time of an SRAM cell (the time required to develop a pre-specified bit-line differential while reading the cell) is a primary component (~40%) of the total memory read time. To improve the quality of an SRAM cell, its leakage and access time needs to be reduced. Hence, we define the quality factor of an SRAM cell as:

$$F=\frac{1}{T_{ACCESS}^{\ \alpha}\cdot I_{leak}^{\ \beta}} \tag{1}$$

where α and β are prespecified parameter based on design consideration. For example, for a high performance SRAM cell, the design is more focused on the access time. Therefore, α is specified to be larger than β. With α and β specified, the SRAM should be designed so that the quality factor given in (1) is maximized.

Due to the regularity of SRAM circuit in the standby mode, the leakage current can be accurately modeled as the summation of the corresponding leakage components. Fig. 2a shows a conventional 6 transistor SRAM cell and the major leakage components during standby mode. Considering the different leakage components of all the transistors, the total leakage of the cell is given by:

$$\begin{aligned} I_{sub} &= I_{subAXR} + I_{subNL} + I_{subPR} \\ I_{jn} &= 2I_{jnAXL} + I_{jnAXR} + I_{jnNL} + I_{jnPR} \\ I_{gate} &= 2I_{gOFF\_AXL} + I_{gOFF\_AXR} + I_{gON\_PR} \\ &\quad + I_{gON\_NR} + 2I_{gOFF\_PL} + I_{gOFF\_NL} \\ I_{leak} &= I_{sub} + I_{jn} + I_{gate} \end{aligned} \tag{2}$$

where, the subscripts associated with the different leakage components represent different transistors within a SRAM cell. In equation (2), the gate leakage is catergorized as the gate leakage in the ON state (for an NMOS the gate terminal is '1' all the other terminals are grounded) and the gate leakage in the OFF state (for an NMOS the drain terminal is '1' and all the other terminals are grounded). The gate leakage in the OFF state consists of only the drain to gate tunneling leakage through the overlap region (the reverse gate leakage). The gate leakage in the ON state includes drain/source to gate overlap leakage, gate to channel leakage and the gate to body leakage. Therefore, the ON state gate leakage is normally an order higher than the OFF state gate leakage. Since, in the SRAM configuration, there are two transistors experiencing the ON state gate leakage during the standby mode, therefore, we have distinguished the gate leakage between the ON state and the OFF state so as to get more accurate results.

It is noticed that during the SRAM access, one of the bit lines (BR in Fig. 2a) is discharged by the ON current through the access transistor (AXR). In the worst case, the other bit line (BL in Fig. 2a) is also discharged by the leakage currents from all the other cells within the same column. Therefore, the access time of the cell can be defined as:

$$T_{ACCESS} = (C_{INT} + NC_{JN})\Delta_{MIN} \Big/ \left(I_{dsatAXR} - \sum_{i=1,..,N} I_{subAXL(i)} \right) \tag{3}$$

where, C_{INT} is the bit-line interconnect capacitance, N is the number of cells in one column of the SRAM array, C_{JN} is the junction capacitance of the access transistors, Δ_{MIN} is the minimum differential that needs to be developed on bitlines for proper reading, I_{subAXL} is the

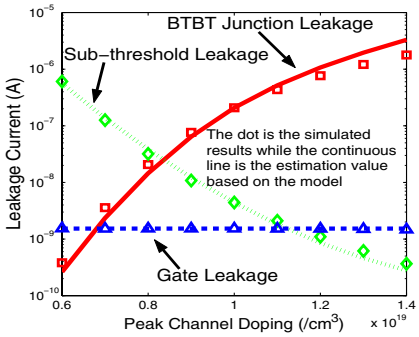


Fig. 4: Doping dependence of different leakage currents

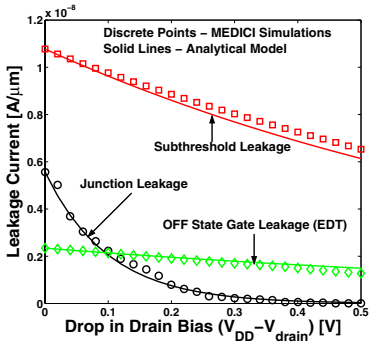


Fig. 5: Voltage swing dependence of different leakage currents

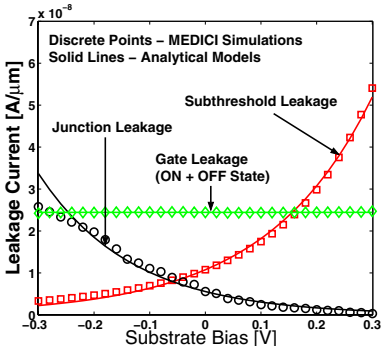


Fig. 6: Substrate bias dependence of different leakage currents

sub-threshold leakage of the acess transistors (AXL) and $I_{dsatAXR}$ is the ON current, saturation current, through the access transistor (AXR in Fig. 2a), which is the current through a stack of two "ON" transistor, access and pull-down NMOS). From (3), it can be observed that to minimize the access time the junction capacitance and the sub-threshold leakage needs to be minimized while the read current (i.e. ON current of the access transistor) needs to be maximized.

3. Device Design and Its Impact on Circuit

From step one (Section 2) of the proposed design methodology, the design specification is modeled by the circuit properties, such as current and capacitance values. In the next step, the circuit properties are related to device design parameters, such as the doping profile.

Conventionally, nano-scale bulk-CMOS devices are optimized using transistor geometry (e.g. oxide thickness) and doping profile. To simplify the problem, in this paper we considered only the doping profile as the device design parameter. It should be noted that the proposed methodology can be extended to take the device geometry in the same way as another design variable for optimization. In scaled technology, device doping is primarily controlled by the "halo" doping concentration. We have used the peak of the "halo" doping to modify the doping concentration in devices [5]. The value of the halo doping modifies different components of leakage current, "ON" current, and junction capacitances. Also the device designed here meets all the ITRS specifications for the 65nm technology [2].

3.1 Leakage in a Nano-scale Si Device

In nano-scale devices, the major components of leakage current are the sub-threshold leakage, the gate leakage and the junction tunneling leakage (Fig. 1). These leakage mechanisms are well studied in previous literature [5] [6]. However, the models developed are too complicated for the proposed *Circuit-aware Device Design Methodology* to get an analytical solution for the optimal device design. Therefore, based on the device physical properties, simple empirical models for the different leakage current versus peaking doping concentration are developed and verified. They are summarized in Table-II in the Appendix. Also, the models are developed below to show the bias dependence of each leakage component.

- Sub-threshold Leakage

The sub-threshold current in a transistor depends exponentially on the gate-to-source voltage and the threshold voltage of a transistor. Due to short channel effect, the sub-threshold current increases with an increase in the drain bias (Drain Induced Barrier Lowering) and a reduction of the channel length (Vth-roll off). Due to the body effect, the sub-threshold leakage reduces with the application of the reverse body-bias. To predict the leakage variation with the terminal voltages, the sub-threshold leakage is modeled as [5]:

$$I_{sub} = I_{sub0} exp\left(\frac{V_{gs} - \eta_{DIBL}\left(V_{DD} - V_{bs}\right) + \lambda_{body} V_{bs}}{m\, kT/q}\right) \quad (4)$$

where, I_{sub0} is the sub-threshold current of a transistors at V_{gs}=0, V_{ds}=V_{DD} and V_{bs}=0, η_{DIBL} is the DIBL coefficient, λ_{body} is the body-effect coefficient and m is the sub-threshold swing factor. Using the values of I_{sub0}, m, η_{DIBL} and λ_{body} which can be extracted from device simulations, the proposed model closely follows the MEDICI [7] simulation results for different gate, drain, and body voltages (as shown in Fig. 5 and Fig. 6).

- Gate direct tunneling leakage

The gate leakage in transistors with ultra-thin gate oxide is due to the direct tunneling of electrons (or holes) through the gate dielectric. Gate leakage increases exponentially with reduction in the oxide thickness and increase in the electric field across the oxide (i.e. oxide voltage). Major components of gate tunneling in a MOSFET are gate to S/D overlap region current, gate to channel current, and gate to substrate current [5]. The overlap current dominates the gate leakage in the "OFF" state whereas gate-to-channel leakage dominates the "ON" state gate leakage. To predict the dependence of the gate leakage on the terminal voltages we model the gate leakage current as:

$$I_{gOFF} = I_{gOFF0} e^{-\alpha_{gOFF}\left(V_{DD} - |V_{gd}|\right)}$$
$$I_{gON} = I_{gON0}\left[e^{-\alpha_{gON}\left(V_{DD} - |V_{gd}|\right)} + e^{-\alpha_{gON}\left(V_{DD} - |V_{gs}|\right)}\right] \quad (5)$$

where, I_{gOFF0} is the OFF state overlap tunneling leakage at $|V_{gd}|$=V_{DD} and I_{gON0} is the ON state gate-to-source leakage at $|V_{gs}|$=V_{DD} (which is equal to the ON state gate-to-drain leakage at $|V_{gd}|$=V_{DD}). To simplify the analysis we also considered that in the "ON" state I_{gs}=I_{gd}=0.5I_{gON} for Vds < 100mV. Using the parameters I_{gOFF0}, I_{gON0}, α_{gOFF}, α_{gON} extracted from device simulations, the proposed model closely follows the MEDICI [7] device simulation.

- Junction tunneling leakage

Junction tunneling leakage is due to the band-to-band-tunneling of electrons in a highly doped reverse biased p-n junction. In nano-scale MOSFETs, due to the use of high junction doping, large junction BTBT occurs with drain at VDD and substrate at ground (at high drain-to-substrate reverse bias) [6]. The junction BTBT exponentially increases with an increase in the drain-to-substrate bias [6]. We model the bias dependence of the junction leakage as:

$$I_{jn} = I_{jn0} exp\left(-\beta_{JN}\left(V_{DD} - |V_{db}|\right)\right) \quad (6)$$

where, I_{jn0} is the junction leakage at $|V_{db}|$=V_{DD} and β_{JN} is a doping dependent empirical fitting factor. Using the extracted values of I_{jn0} and β_{JN} from the device simulations the above model closely matches the MEDICI simulations.

In Appendix (Table-II), empirical models are developed for the above-mentioned leakage currents in terms of the peak halo doping concentration. Fig. 4 shows the doping dependence of the different leakage components based on the developed empirical models in Table-II and the MEDICI simulations. Increasing the halo doping concentration (i.e. by increasing the peak halo doping vaule) reduces the sub-threshold leakage (less DIBL and Vt roll-off effects). However, a higher halo doping increases the electric field across the junction resulting in a higher junction tunneling current. Gate leakage

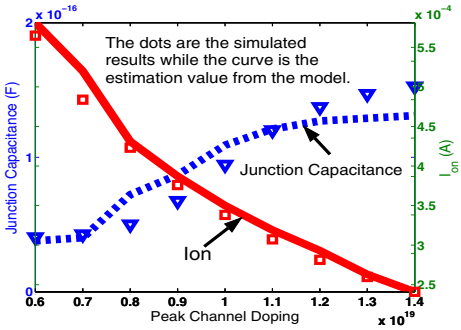


Fig. 7: Doping dependence of Ion and junction capacitance

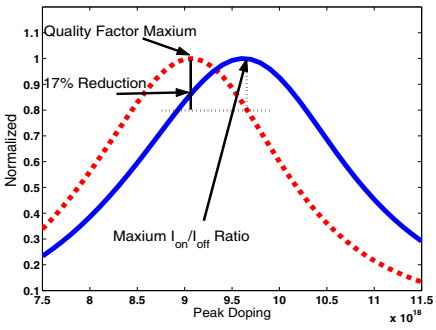


Fig. 8: Doping dependence of 6T-SRAM quality factor and the doping dependence of the device *Ion/Ioff* ratio (normalized)

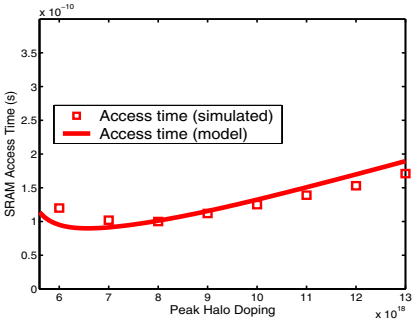


Fig. 9: Device doping dependence of the SRAM cell access time

is only weakly sensitive to any change in the doping. Hence, modification of the peak halo doping is an effective method for optimizing devices for leakage current. Also from Fig. 4, it is shown that the developed empirical models (Table-II) correctly predict the doping dependence of the different leakage components.

To verify the bias dependence of the developed leakage models (equation 4-6), the leakage currents under different bias condition are simulated and compared with the developed models. Fig. 5 shows the value of the different leakage components versus the voltage swing across the drain and source terminals of a transistor. In the simulation, the source and gate voltage (NMOS) are fixed at GND while the drain voltage sweeps from VDD to VDD-0.5 (V). As shown in Fig. 5, the developed models correctly track the MEDICI simulation results. Also, as the voltage swing reduces, all the leakage components reduce either linearly (sub-threshold leakage, gate leakage) or exponentially (junction leakage). This effect is similar to the power saving mechanism in the case of SB-SRAMs. In Fig. 6, the leakage components under different body biasing voltage is shown. Again, the developed models correctly predict the leakage current. Also, the negative biasing reduces the sub-threshold leakage. However, the junction leakage is higher due to the increase of voltage across the body/source and body/drain junction in the negative biasing mode. These phenomena are also observed in the reverse biased SRAM (RB-SRAM).

3.2 The ON Current and the Junction Capacitance

An empirical Ion versus peak doping model is also derived in Table-II. As shown in Fig. 7, the empirical model matches the MEDICI simulation results for a wide doping range from 0.6e19 to 1.4e19. Increasing the peak halo doping increases the Vt of the device, resulting in a lower "ON" current (Fig. 7).

On the other hand, the performance of a circuit also depends on the parasitic capacitances associated with drain-substrate and source-substrate junctions. The simplified doping dependence model (Table-II) for the junction capacitance is developed and plotted in Fig. 7 together with the results from MEDICI simulation. It is shown in Fig. 7 that an increase in the junction doping reduces the depletion width across the drain-substrate and source-substrate junctions resulting in a higher junction capacitance.

It should be noted that after the development of the above-mentioned doping dependence models, all the circuit properties are simple functions of the device design variables (doping in this case). Based on the models, it is shown that a reduction in the peak doping in the transistors reduces the access time due to higher read current and lower junction capacitance. However, reducing doping too much has an adversary effect on the access time due to an expeonential increase in the sub-threshold leakage. Also, a reduction in the peak doping increases the sub-threshold leakage contribution to the cell, whereas increasing the doping increases the contribution of the junction leakage. Therefore, using (equation 2) and (equation 3) the cell quality factor can be represented in terms of the device design paramters (doping concentration in this case).

4. Circuit-aware Device Optimization

Based on the models developed in the first step (Section 2) and the second step (Section 3) of the proposed *Circuit-aware Device Design* methodology, the SRAM design quality factor is presented as a function of the device design parameters (doping). Therefore, the optimal device, which maximizes the quality factor, can be selected analytically.

In this section, we investigate the effects of *Circuit-aware Device Design* methodology for designing conventional 6-T SRAM, SB-SRAM and BB-SRAM.

4.1 Conventional 6-T SRAM

Following the *Circuit-aware Device Design* methodology shown in Fig. 3b, the optimal device can be chosen by taking the derivative of the quality factor versus the doping concentration. Therefore, an analytical form of the optimal doping concentration can be derived.

Fig. 8 shows the doping dependence of the quality factor ($\alpha = \beta = 1$) of a 6-T SRAM. Together with the quality factor, the doping dependence of the Ion/Ioff ratio of a NMOS is also plotted. It should be noted that the plots here are normalized with respect to their own maximum values. It is observed that, the device with maximum Ion/Ioff value has 17% reduction in the SRAM quality factor. Also, to show the advantage of the proposed methodology more intuitively, circuit level MEDICI simulation is performed to evaluate the SRAM access time and leakage. Fig. 9 shows the device doping dependence of the SRAM access time from MEDICI simulations, which also verifies equation 3. It is observed that the optimal device from the proposed methodology reduces the SRAM access time by 6.6% compared with the max Ion/Ioff ratio device. Fig. 10 plots the device doping dependence of the SRAM standby leakage current from the MEDICI simulations and from our model (equation 2). The proposed *Circuit-aware Device Design* methodology also achieves an SRAM with 10.7% reduction in leakage.

In Fig. 11, the 6-T SRAM quality factor curves for different α and β values are plotted. It is shown in Fig. 11 that the optimal devices for different design purposes are different. Especially, the optimal device for high performance SRAM designs has the peak doping considerably lower than the optimal device for low power SRAM designs. Therefore, the device design should be adjusted according to the design applications.

4.2 Source-Biasing SRAM (SB-SRAM)

In SB-SRAM (Fig. 2c), a positive source bias voltage is applied in the standby mode to reduce cell leakage. In the active mode, the bias is removed to maintain performance. In case of SB-SRAM, the cell leakage current model (equation 2) is modified (using equation 4-6) to include the effect of source bias on different leakage components. Moreover, the read current ($I_{dsatAXR}$ in equation 3) needs to be determined considering the discharge current through a 3 transistor stack, instead of a 2 transistor stack in a conventional SRAM, due to the fact that normally a sleep transistor is inserted between the virtual ground and the real ground in SB-SRAM. Considering a source bias (V_{SL}) of 0.3V, we optimize the doping profile of the devices using the

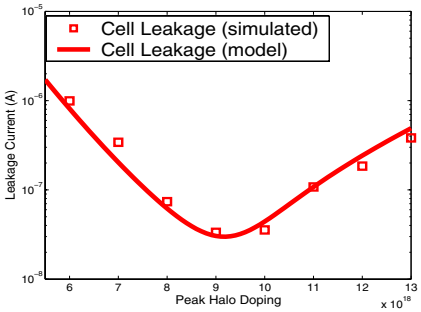


Fig. 10: Device doping dependence of the overall SRAM cell leakage

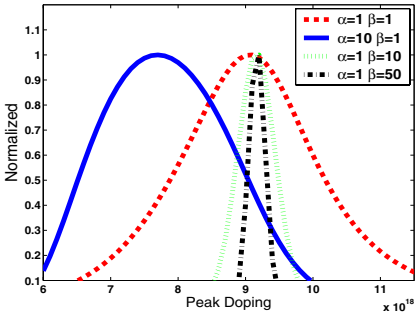


Fig. 11: Quality factor curve for different SRAM design purpose

Table-I: Optimal device under different design methodology

Design Criteria	Doping	I_{on} (A/um)	I_{leak} (A/um)
Max(Ion/Ioff)	9.60E+18	6.60E-04	3.75E-08
Proposed (6T SRAM)	9.00E+18	6.97E-04	4.51E-08
Proposed (SB-SRAM)	5.10E+18	1.20E-03	5.97E-06
Proposed (RBB-SRAM)	8.60E+18	7.37E-04	6.52E-08

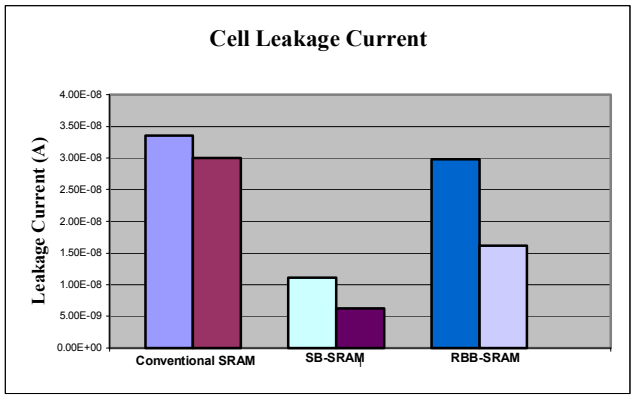


Fig. 12: Cell leakage with different device designed methodologies

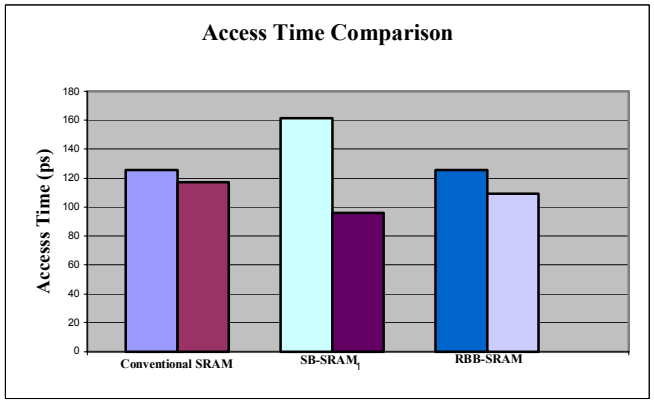


Fig. 13: Cell access time with different device designed methodologies

defined quality factor function, with the modifications required for the source bias. With source bias, the NMOS AXR and NL (Fig. 2c) have negative Vgs (=-V_{SL}), negative Vbs (=-V_{SL}) and lower Vds (=VDD-V_{SL}). Hence, sub-threshold leakage reduces significantly. Source bias also reduces the ON state gate leakage of transistor NR. Hence, the total cell leakage is dominated by the junction tunneling leakage, which is only weakly sensitive to source bias (increases marginally due to the small leakage through the source-substrate junctions of NL, NR, AXR). Due to this fact, for SB-SRAM, the optimal peak doping of the devices can be considerably low, compared to the max Ion/Ioff ratio criteria for device design, so as to reduce the transistor junction leakage as well as the total SRAM cell leakage. The reduction in the doping also improves the access time. However, considerable reduction of doping also adversely affects the access time due to the increase in the sub-threshold leakage. Increase in the DIBL effect also limits the peak doping reduction in the devices for SB-SRAM. Based on our proposed *Circuit-aware Device Design,* the optimum doping for the device of SB-SRAM is 5.1e18, which is significantly lower compared to that of the max Ion/Ioff device (see Table-I). The optimal device results in 44% reduction in cell leakage (Fig. 12) and 40% improvement in cell access time (Fig. 13) for the SB-SRAM cell compared to the SB-SRAM designed with the max *Ion/Ioff* device.

4.3 Body-Biased SRAM (BB-SRAM)

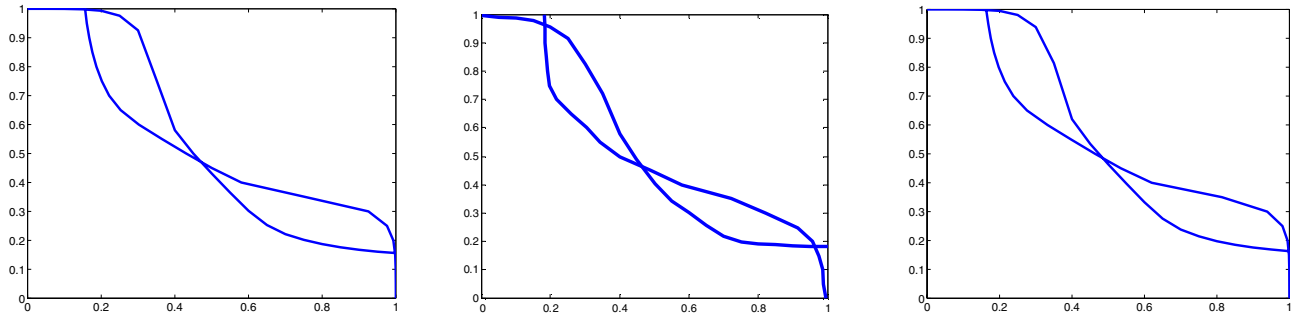
In body-biased SRAM, a different voltage bias is applied to the body of the devices in the stand-by mode (to reduce leakage) or in the active mode (to improve performance). There are principally two types of BB-SRAM's: (a) Reverse-Body-Bias SRAM (RBB-SRAM) where, RBB is applied in the stand-by mode and zero body bias (ZBB) is applied in the active mode, and (b) Forward-Body-Bias SRAM where FBB is applied in the active mode and ZBB is applied in the stand-by mode. In this work, we focus on the RBB-SRAM because forward body bias is shown to be less effective in scaled technology.

The cell leakage current (equation 2) is modified to include the effect of RBB on the leakage currents (assuming -0.25V of reverse bias in the stand-by mode). The read current ($I_{dsatAXR}$ in equation 3) remains the same, as no body bias is applied in the active mode. With RBB, sub-threshold leakage reduces and junction leakage increases. Gate leakage is insensitive to the body-bias. Hence, reducing the peak halo doping, which reduces junction leakage, helps to reduce the overall cell leakage in the RBB-SRAM. The optimized device for the RBB-SRAM cell also has a lower peak halo doping compared to the max *Ion/Ioff* ratio device (Table-I). The RBB-SRAM cell from the *Circuit-aware Device Design* flow has 43.2% lower leakage and 12.7% better cell access time compared with the RBB-SRAM designed using max *Ion/Ioff* ratio device, as shown in Fig. 12 and Fig. 13.

4.4 Observations

Table-I listed the optimal device under different design criteria (*IndOptDevCirc-design* and *Circuit-aware Device Design* for conventional 6T-SRAM, SB-SRAM and RBB-SRAM). For all the SRAM styles, the proposed device design methodology results in lower doping values compared with the max Ion/Ioff criteria for device design. With the reduction of the peak doping, the Ion current increases (Table-I), which results in less cell access time. On the other hand, the lower doping values increase the individual transistor leakage current (I_{leak} in Table-I) due to the exponential dependence of sub-threshold leakage on doping.

Fig. 12 shows the SRAM cell leakage current for conventional 6T-SRAM, SB-SRAM, and RBB-SRAM, respectively. The left bar of each circuit style shows the standby leakage of SRAM cells designed with the max *Ion/Ioff* ratio transistor. The right bar is the SRAM cell leakage with the optimal device based on the proposed *Circuit-aware Device Design* methodology. Similarly, Fig. 13 shows the cell access time of three SRAM circuit styles with devices designed under the *IndOptDevCirc-design* flow and the proposed design flow. The proposed methodology reduces the cell access time for all three SRAM styles. This is due to the increase of device Ion shown in Table-I. More importantly, though the proposed methodology increases the individual transistor leakage current (I_{leak} in Table-I), it reduces the SRAM cell leakage as shown in Fig. 12. This is due to the fact that within SRAM circuits, there are more transistors experiencing junction leakage component than the number of



(a) Conventional 6T-SRAM (b) SB-SRAM (c) RBB-SRAM

Fig 14: Read SNM for the SRAM design with the optimal device

transistors with sub-threshold leakage components (equation 2). Therefore, taking advantages of this circuit property, a lower doping device, which reduces the individual transistor junction leakage, also results in an overall reduction of SRAM leakage. Furthermore, the circuit techniques (source biasing or body biasing) are applied mainly to reduce sub-threshold leakage as it used to be the dominant leakage component in technologies above 100nm. Therefore, the optimal device doping for SB-SRAM and RBB-SRAM can be further reduced so as to balance the sub-threshold leakage and junction leakage value in the SRAM circuit. The considerable leakage reduction of SB-SRAM and RBB-SRAM through the proposed design flow is observed in Fig. 12.

The major observations from the proposed *Circuit-aware Device Design* methodology can be summarized as:

- Optimization of the transistors individually for maximum Ion/Ioff ratio does not necessarily result in optimum performance and leakage power in SRAM cells.
- The devices, which optimize the power and performance in different SRAM circuit styles, are different from each other. The optimal device is determined by the circuit properties (i.e. which leakage component is reduced by the applied circuit technique).
- The optimal devices from the *Circuit-aware Device Design* methodology significantly improve the leakage and performance of the SRAM designs.

5. Static Noise Margin

In SRAM design, besides leakage power and access time, there are other design specifications, such as SNM. In Fig. 14, the butterfly curves during the cell read operation of a conventional 6-T SRAM, a SB-SRAM and a RBB-SRAM are plotted. All these SRAM designs have used the optimal device based on the proposed methodology. As shown in Fig. 14, the proposed design methodology maintains reasonable read SNM for all the SRAM designs. It should be noted that the SNM can be also included in the quality factor (equation 1) of the SRAM cell design. Or the SNM can be used as a constraint for SRAM design. A simple SNM model was developed in [8]. Using this model, we can take the SNM as an optimization goal or a constraint during the proposed methodology for SRAM design. Following the same flow, the optimal device, in terms of the new quality factor, can be achieved analytically.

6. Conclusions

In this paper we proposed a general *Circuit-aware Device Design* methodology, which analytically derives the optimal device design in terms of the pre-specified circuit design and the quality factor. We applied the methodology to SRAM design and achieved up to 44% of leakage reduction and 40% of access time improvement (in SB-SRAM) compared to the SB-SRAM designed with the max *Ion/Ioff* device. We also observed that the optimal devices for different circuit techniques are different and the optimal device depends considerably on the circuit properties. Therefore, to further improve the performance and power of a digital design, device design and circuit design should not be isolated, especially for scaled technologies. We believe the proposed *Circuit-aware Device Design* methodology will be very useful for sub-90nm technology where all the leakage components are comparable in magnitude.

7. Appendix:

The developed empirical models for the doping dependence of the device properties (such as leakage currents, Ion and junction capacitance), described in section 3, are summarized in Table-II. In these equations, all the parameters with the subscript "Ref" represents the values measured at a reference peak halo doping concentration. N_{cheff} is the peak halo doping of the device. N_{cheff_Ref} is the reference peak doping. All α, β, γ and m values can be extracted from device simulations for a particular device structure. All these simplified equations are verified as shown in Fig. 4 and Fig. 7.

Acknowledgement

This work is sponsored in part by MARCO GSRC and National Science Foundation (NSF).

Reference:

[1] Mukhopadhyay, S., et. al., "Leakage in nanometer scale CMOS circuits", International Symposium on VLSI Technology, Systems, and Applications, Oct. 2003.
[2] The National Roadmap for Semiconductors, S.I.A., 2000.
[3] Kim, C.H., et. al., "A forward body-biased low-leakage SRAM cache: device, circuit and architecture considerations", TVLSI, Volume 13, Issue 3, March 2005.
[4] Mukhopadhyay, S., et. al., "Accurate estimation of total leakage current in scaled CMOS logic circuits based on compact current modeling", DAC, 2003.
[5] Roy, K., et. al., "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits", IEEE Proceedings, Volume 91, Issue 2, Feb. 2003.
[6] Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices.
[7] MEDICI: Two-dimensional semiconductor device simulation program, AVANT! Corp., Fremont, CA, 2000.
[8] Bhavnagarwala, A., et. al., "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," IEEE JSSC, pp. 658-665, April 2001.

Table-II: Empirical models for the doping dependence of leakage currents, Ion and junction capacitance

Sub-threshold Leakage	$I_{sub} = I_{sub\_Ref} \exp[\alpha(\frac{N_{cheff}}{N_{cheff\_Ref}})^2 + \beta(\frac{N_{cheff}}{N_{cheff\_Ref}}) + \gamma]$
Junction Leakage	$I_{jn} = I_{jn\_Ref} \cdot \exp[\frac{\alpha(N_{cheff} - N_{cheff\_Ref})}{\sqrt{N_{cheff}}}]$
Gate Leakage	$I_{gate} = I_{gate\_Ref}$
Ion	$I_{on} = I_{on\_Ref} \frac{N_{cheff\_Ref}}{N_{cheff}}$
Junction Capacitance	$C_j = C_{Ref} \cdot (\frac{\frac{N_{cheff}}{N_{cheff\_Ref}}}{\alpha + \beta \frac{N_{cheff}}{N_{cheff\_Ref}}})^{\gamma} \cdot \phi(N_{cheff})$ $\phi(N_{cheff}) = \frac{2(\frac{N_{cheff}}{N_{cheff\_Ref}})^m}{1 + (\frac{N_{cheff}}{N_{cheff\_Ref}})^m}$