

Power/Performance Based Scalability Comparisons between Conventional and Novel Transistors Down to 32nm Technology Node

P. Kapur, R. S. Shenoy, and K. C. Saraswat

Center for Integrated Systems (CIS) 110

Stanford University

Stanford, CA 94305 U.S.A

kapurp@stanford.edu

Abstract -- We quantify and compare the scalability of bulk, partially depleted SOI, and double gate transistors with and without high-k gate dielectric down to 32nm technology node in terms of globally optimized power/performance curves. The novelty of work is in that it provides a quantitative tool to determine the suitable insertion point for novel transistor schemes. It also addresses optimum supply/threshold voltage, gate dielectric thickness, and doping concentration scaling, unique to different devices and circuit functional blocks.

I. INTRODUCTION

Scaling-induced dramatic rise in leakage power [1] has prompted aggressive search for solutions, which mitigate this problem at different levels including architecture, circuits and devices. In the area of devices, this has led to an explosion of novel ideas in the structural (e.g. multi-gate FETs) and materials domain (e.g. high-k gate dielectric). A fair comparison of the efficacy of these solutions at future nodes, and their advantages over the currently prevalent Bulk/SiO₂ gate transistor requires a comprehensive comparison methodology. In this work, we develop this standard using globally optimized power/performance curves. A unique power/performance curve for each type of device is obtained by a multi-dimensional optimization of supply and threshold voltage (V_{dd} , V_t), doping concentration (N_a), and equivalent oxide thickness (EOT) for gate dielectric. The methodology can serve as a powerful tool for device and circuits community by aiding in 1) Device selection at future nodes with no SPICE models, 2) Optimum device design once appropriate device is selected, 3) predicting V_{dd} scaling trends for different functional blocks depending on the device selection.

In our previous work, we showed a limited application of this methodology to double gate transistors (DGFET) [2]. In this work, we expand the scope to compare six different innovative transistor schemes consisting of bulk, partially depleted (PD) SOI, and DGFET with either high-k or with SiO₂ gate dielectrics (Fig. 1). To study the impact of scaling, we consider two gate lengths (L_g) of 18nm and 14nm, targeting 45 and the 32nm high performance node [3]. Because the purpose of this work is to show comparison trends only, we consider uniformly doped bulk devices (no halo or V_t implants), representing their worst-case. The optimized global advantage arising from removal of junction capacitances (C_{diff}) is considered. This is henceforth named as the PDSOI device. Although, no floating body effects are considered. Further, the DGFET power is that of a single gate and all devices use metal gate work-function (ϕ_m) to set appropriate V_{ts} .

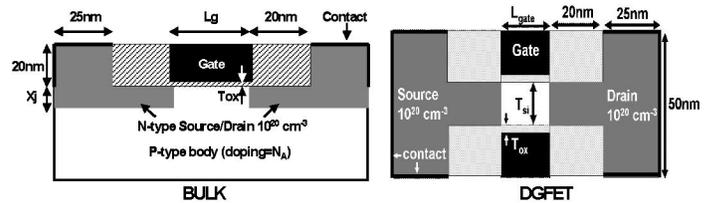


Fig. 1: Schematic of Medici-simulated bulk and double-gate (DG) FET. Both 45nm and 32nm HP L_g chosen according to ITRS 2003 (45nm \Rightarrow 18nm, 32nm \Rightarrow 14nm).

II. METHODOLOGY

For a given FO1 inverter delay of a given transistor type, we minimize the sum of dynamic (DP), sub-threshold (SDL) and gate leakage powers (GL) by optimizing V_{dd} , V_t , N_a (bulk/PDSOI only), and EOT. A local power minimum with V_{dd} (implicitly V_t) and a global one with respect to EOT is typical and is shown for a sample FO1 delay in Fig. 2. The V_{dd} optimum is a result of apposite trends in various power components with respect to V_{dd} . For example, SDL reduces with V_{dd} because a constant delay condition allows a higher V_t at a higher V_{dd} . Whereas, DP and GL, as expected, increase with V_{dd} . The EOT optimum, on the other hand, is a result of a balance between GL and SDL. This global optimization is repeated for different delays to generate the optimized power-delay (performance) curve for a particular transistor. DP, SDL and GL calculations required extensive device simulations (I-V, C-V curves), which served as the input to this methodology [2]. Existing gate leakage models and analytical models for DGFET devices were used [4], [5].

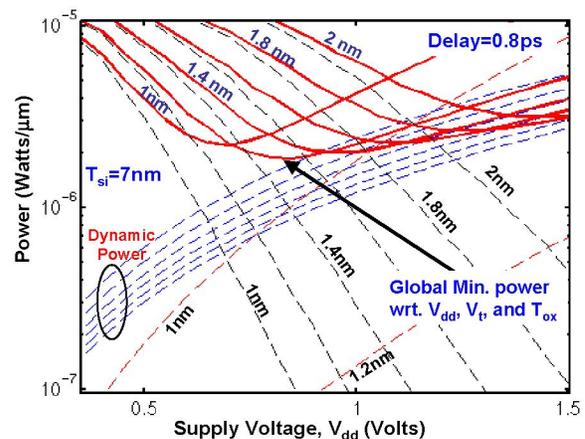


Fig. 2: Total power curves for different EOT (solid lines) for a target delay of 0.8ps ($L_g=18$ nm, S.A=10%, DGFET). Dashed curves show dynamic (DP), S/D leakage (SDL), gate leakage (GL). GL and DP rise with V_{dd} , SDL falls.

III. RESULTS

A. EOT Optimization, Comparison and Scaling

With SiO₂ gate dielectric, at a fixed delay, we observe an optimum EOT minimizing total power (Fig. 3, Fig. 4). As discussed earlier, the optimum is a consequence of the tradeoff between GL and SDL. Fig. 3 shows that the optimum EOT increases with doping concentration for a given bulk device. This is because of a lower V_t and a resulting higher vertical field for a higher N_a devices. Fig. 3 also clearly depicts a more dramatic relief in power with EOT reduction for lower N_a owing to a worse short channel control in lower N_a devices. Fig. 4 compares the EOT optimization for different types of transistors. The presence of the extra C_{diff} in the bulk transistor compared to PDSOI, requires a lower V_t to get the same delay, increasing GL, thus, forcing a larger optimum EOT. Further, the DGFETs have low vertical electric field (hence, GL), thus, can afford a low EOT. In addition to the optimum EOT trend, Fig. 4 also shows the bulk devices exhibiting the most and the DGFETs (already good SCE) the least improvement in the power with EOT reduction (using high-k). This is because the existence of the parasitic C_{diff} in bulk devices affords an additional increase in V_t with EOT reduction on top of that gained because of improvement in SCE. Thus, high-k dielectrics are most advantageous for bulk devices.

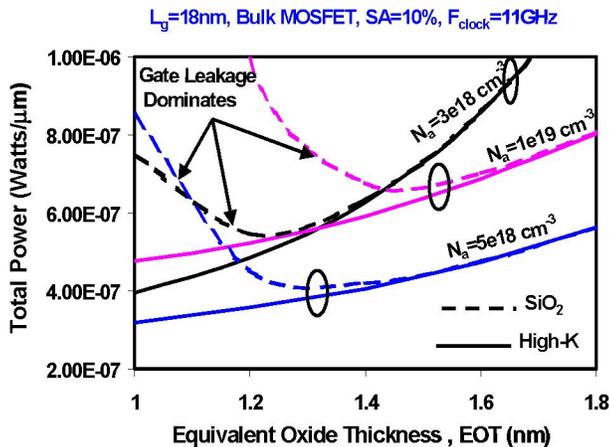


Fig. 3: Figure showing existence of optimum EOT and its trends with different doping concentration for bulk MOSFET

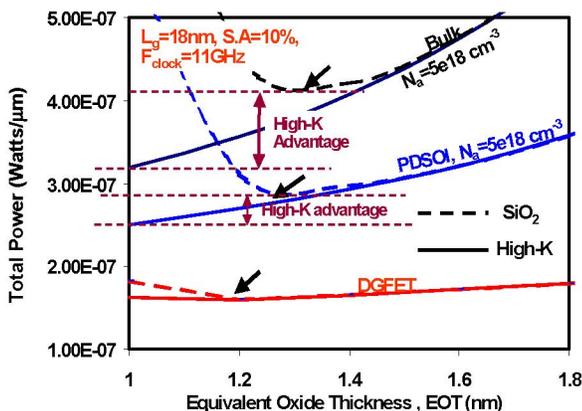


Fig. 4: Variation in optimum EOT as a function of different devices. Arrows point to optimum EOT for the SiO₂ gate dielectric. High-K advantage is largest of bulk.

Finally, optimum EOT (for SiO₂ gate dielectric) as a function of technology scaling is studied by decoupling the impacts of the accompanying delay, and L_g reduction. A decrease of ~30% in both L_g and delay results in ~10% increase in optimum EOT. This is primarily because of a higher GL at faster speeds due to a lower V_t requirement.

B. Optimized Power/Performance Curves and Doping

Fig 5 plots the globally optimized power/performance curves for three devices, Bulk, PDSOI, and DGFET with high-k gate dielectric and L_g of 18nm (45nm node). Both bulk and PDSOI, in turn, have three curves corresponding to different doping concentrations. Fig. 6 shows a similar plot for an L_g of 14nm, roughly corresponding to 32nm technology node. In these figures, the DGFET dissipates the least power followed by PDSOI and bulk, with the discrepancy increasing for higher F_{clock}. Within the bulk or PDSOI devices there is an indication of an optimum with respect to doping. This is explicitly seen in Fig. 7, which plots the optimum total power vs. N_a at different F_{clocks}. The optimum N_a is a result of the tradeoff between better electrostatics, but worse mobility, subthreshold slope, and C_{diff} as N_a is increased. Fig. 7 also shows that the optimum N_a for PDSOI is larger than bulk. This is because an increase in N_a results in an extra penalty for bulk compared to PDSOI in the form of an increases in C_{diff}.

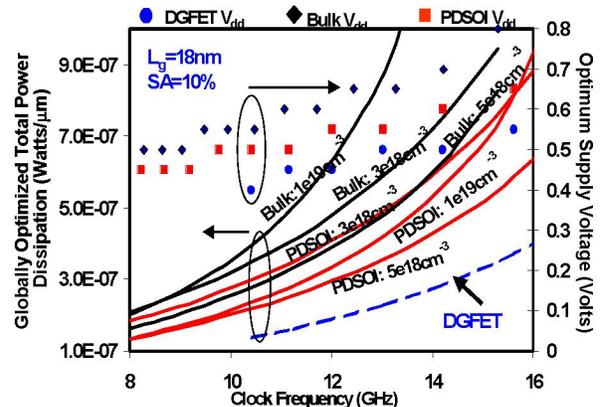


Fig. 5: The globally optimized power vs. clock frequency curves (left axis) for competing devices (Bulk, PDSOI and DGFET) with high-k at 45nm HP node. The dots show corresponding optimum V_{dd} (only 5e18 cm⁻³ doping is plotted for bulk/PDSOI) on right axis.

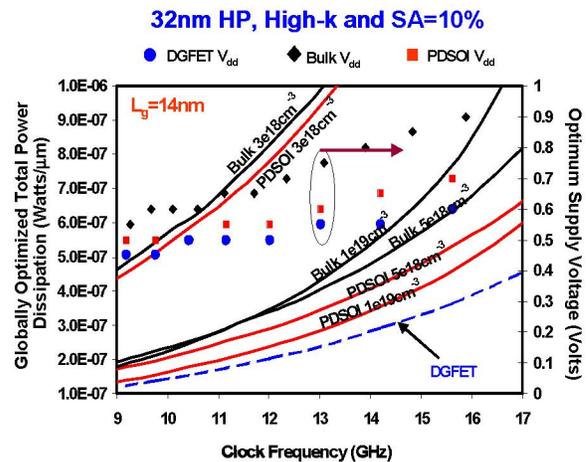


Fig. 6: The globally optimized power vs. clock frequency curves (left axis) similar to Fig. 5 but for the more aggressive 32nm high performance node.

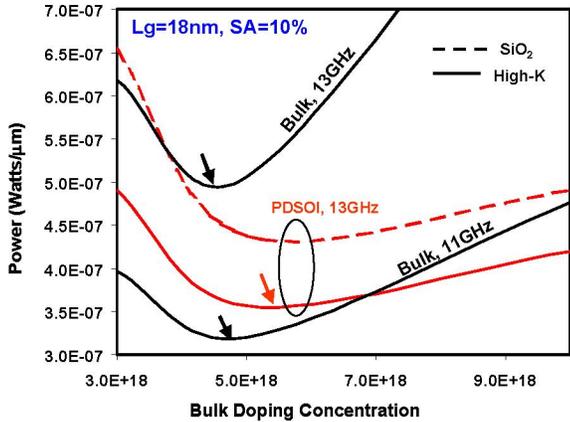


Fig. 7: Total Power showing an optimum with respect to doping concentration. (tradeoff between SCE improvement and mobility degradation). Shown are impact of i) high-k vs. SiO₂, ii) PDSOI vs. bulk, and iii) clock frequency on optimum doping

Fig. 8 shows the impact of scaling on optimum N_a for bulk devices. Reduction in L_g and increase in F_{clock} are decoupled. A smaller L_g clearly shows a larger optimum N_a . This is because a lower L_g has worse SCE, thus, requiring a larger N_a to balance this.

C. V_{dd} Optimization, Comparison and Scaling

Optimum V_{dd} corresponding to the best power/performance curves is dependent on the type of device. As can be seen from Fig. 5 (right axis), that the bulk devices yield the largest values for optimum V_{dd} (0.6–0.8V). This is because at optimum V_{dd} , SDL is universally $\sim 20\text{-}30\%$ of DP and bulk with its higher SDL meets above condition at a higher V_{dd} . We also find that the SiO₂ based devices exhibit a higher optimum V_{dd} compared with high-k devices (not shown here) owing to their larger optimum EOT value. A larger EOT has a higher SDL and a lower DP, thus requires a larger V_{dd} before SDL becomes $\sim 20\text{-}30\%$ of DP. Technology scaling involves L_g reduction and a possible F_{clock} increase. Higher F_{clock} exhibits a higher optimum V_{dd} as seen in Fig. 5, whereas, optimum V_{dd} remains relatively unchanged with L_g reduction at the same F_{clock} .

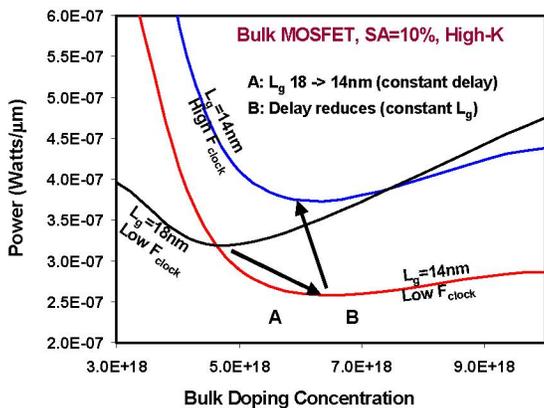


Fig. 8: Impact of scaling on the optimum doping concentration of the Bulk transistor. Scaling is decoupled as L_g reduction and F_{clock} increase

D. Scaling Requirements for Different functional blocks

Different functional blocks (registers, data paths, and clocks) with their unique switching activities (SA) require different transistor designs for best performance. We consider three different SAs of 1%, 10%, and 50% corresponding approximately to registers, data paths and clocks, respectively. We find that a higher SA circuit requires a 1) lower optimum V_{dd} , 2) lower optimum T_{ox} with SiO₂ gate dielectric, and 3) a lower optimum N_a . A higher SA circuit has a proportionately larger DP. Thus, needs a lower V_{dd} before SDL drops to 20–30% of DP. Whereas, the optimum EOT is lower for higher SA (Fig. 9) because GL is less important at lower V_{dd} (optimum V_{dd} is lower). Fig. 10 shows clearly shows that a higher SA circuit also requires a lower N_a . As discussed before, the optimum V_{dd} is lower for higher SA circuits. Also, the SCE are less pronounced at lower V_{dd} ; thus, there is less incentive to go to higher doping (which only mitigates SCE) for higher SA circuits.

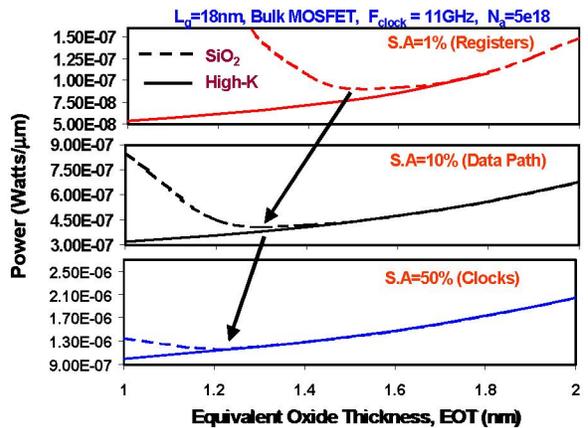


Fig. 9: EOT optimization for different chip functional blocks marked by different switching activity.

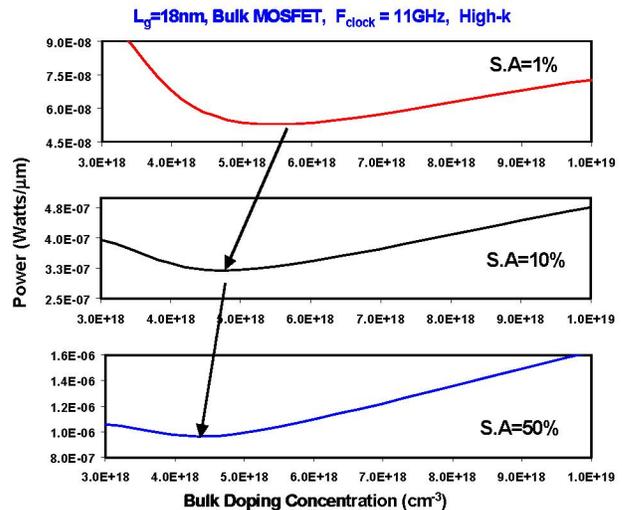


Fig. 10: Optimum doping concentration for different functional blocks on a chip marked by their difference in the switching activity.

IV. DISCUSSION AND SUMMARY

The results in this paper deal either with the power savings by comparing globally optimized versions of different transistors, or with the optimum parameters to achieve those powers. Tables 1 and 2 summarize the optimized power numbers, whereas, the Table 3 summarizes the optimum parameter trends. Table 1 and 2 quantify the power advantage of novel schemes over conventional bulk/SiO₂ transistor for SA of 1% and 10%. Comparing these tables and a similar table for SA=50% (not shown here), we find all innovations to be most effective for low SA circuits, with SA=1%, showing 20%-70% of the power exhibited by the bulk/SiO₂ transistor (Table 1), while SA=10% showing 31%-78% of the bulk/SiO₂ power (Table 2). For a given SA, (e.g. Table 2), when comparing different innovations (down the column), we find several interesting trends: i) C_{diff} removal (bulk vs. PDSOI) yields similar advantage (25% at 9GHz) as introducing high-k (20% at 9GHz). ii) DGFET is the best solution iii) High-k is most effective for bulk (~25% saving bulk, 17% PDSOI, 11GHz, L_g=18nm). Further, the advantage of all innovations increases with F_{clock} and with technology scaling.

Table 1: A table quantifying the power advantage for various novel schemes over standard bulk/SiO₂ transistor. Power is normalized with respect to the top row of each column, which is the Bulk/SiO₂ transistor row. Thus, the column numbers quantify the relative power compared to the current paradigm. Inside parenthesis are actual power number in Watts/μm. This is for a SA of 1%.

		L _g =18nm (45nm node)			L _g =14nm (~32nm node)		
		9GHz	11GHz	13GHz	9GHz	11GHz	13GHz
SiO ₂	Bulk	1 (5.27e-8)	1 (9.23e-8)	1 (1.85e-7)	1 (4.02e-8)	1 (7.19e-8)	1 (1.24e-7)
	PDSOI	0.65 (3.45e-8)	0.6 (5.55e-8)	0.47 (8.66e-8)	0.69 (2.77e-8)	0.58 (4.2e-8)	0.49 (6.13e-8)
	DGFET	0.34 (1.8e-8)	0.3 (2.72e-8)	0.21 (3.885e-8)	0.54 (2.16e-8)	0.43 (3.12e-8)	0.35 (4.36e-8)
High-k	Bulk	0.65 (3.44e-8)	0.58 (5.37e-8)	0.47 (8.76e-8)	0.7 (2.79e-8)	0.63 (4.51e-8)	0.55 (6.85e-8)
	PDSOI	0.54 (2.83e-8)	0.46 (4.25e-8)	0.33 (6.15e-8)	0.56 (2.24e-8)	0.45 (3.23e-8)	0.37 (4.61e-8)
	DGFET	0.33 (1.74e-8)	0.28 (2.58e-8)	0.2 (3.64e-8)	0.49 (1.96e-8)	0.39 (2.78e-8)	0.31 (3.79e-8)

Table 2: Very similar table to Table 1, except this one has a SA of 10%

		L _g =18nm (45nm node)			L _g =14nm (~32nm node)		
		9GHz	11GHz	13GHz	9GHz	11GHz	13GHz
SiO ₂	Bulk	1 (2.58e-7)	1 (4.22e-7)	1 (7.34e-7)	1 (2.24e-7)	1 (3.85e-7)	1 (5.81e-7)
	PDSOI	0.76 (1.97e-7)	0.71 (2.98e-7)	0.6 (4.38e-7)	0.71 (1.6e-7)	0.62 (2.38e-7)	0.59 (3.41e-7)
	DGFET	0.42 (1.08e-7)	0.38 (1.61e-7)	0.32 (2.32e-7)	0.58 (1.3e-7)	0.49 (1.87e-7)	0.44 (2.58e-7)
High-k	Bulk	0.8 (2.06e-7)	0.76 (3.2e-7)	0.69 (5.06e-7)	0.78 (1.74e-7)	0.74 (2.86e-7)	0.71 (4.1e-7)
	PDSOI	0.66 (1.7e-7)	0.59 (2.5e-7)	0.49 (3.56e-7)	0.62 (1.39e-7)	0.53 (2.02e-7)	0.49 (2.85e-7)
	DGFET	0.42 (1.08e-7)	0.38 (1.59e-7)	0.31 (2.28e-7)	0.55 (1.23e-7)	0.45 (1.75e-7)	0.41 (2.4e-7)

Another interesting trend from Tables 1 and 2 can be found by comparing across rows to a different technology node. It is found that at same delay, the lower L_g yields a lower power, making devices more energy efficient with scaling. Although, this advantage is small. Also, increase F_{clock} by 30% for bulk/SiO₂ roughly results in 3X increase in optimized power at the same gate length. This increase is somewhat mitigated by going to a lower L_g.

Table 3 summarizes design rules for device and circuit designers, showing qualitative design requirements under different scenarios involving different: 1. transistors, 2. functional blocks, and 3. scaling (F_{clock} and L_g).

In summary, we have quantified optimum power trends and the corresponding device and circuit parameters (EOT, N_a and V_{dd}) for different devices, for different functional blocks, and as a function of L_g and F_{clock} scaling.

Table 3: A table summarizing the design rules and guidelines for Different types of transistors, for different functional blocks with Different SAs, for different F_{clock}, and finally with technology scaling.

▲ Least ↑ Moderate ↑ Highest

		Gate Dielectric		Optimum Doping	Optimum V _{dd}	
		Optimum EOT (SiO ₂)	High-k Advant.			
Different Structures	Bulk	↑	↑	↑	↑	
	PDSOI	↑	↑	↑	↑	
	DGFET	▲	▲	N/A	▲	
Different S.A.	1%	↑	↑	↑	↑	
	10%	↑	↑	↑	↑	
	50%	▲	▲	▲	▲	
Different F _{clock}	Low	▲	▲	↑	▲	
	High	↑	↑	▲	↑	
Scaling (L _g Reduction)	EOT: reduced				▲	
	F _{clock} : not increased				▲	
	EOT: not reduced				↑	
	F _{clock} : not increased				▲	
					EOT: not reduced	↑
					F _{clock} : increased	↑

REFERENCES

- [1] T. Kuroda, IEICE Trans. Electron., Vol. E84-C, No. 8, Aug 2001, pp. 1021-1028.
- [2] P. Kapur, R.S. Shenoy and K.C. Saraswat, *International Electron Device Meeting (IEDM) technical digest*, 2004.
- [3] *International Technology Roadmap for Semiconductors, ITRS SIA* 2003.
- [4] W. Lee and C. Hu, *IEEE Transaction on Electron Devices*, vol. 48, no.7, pp.1366-73 2001.
- [5] Y. Taur, *IEEE Electron Device Letters*, vol. 21, no.5, 2000, pp. 245-247.