Computing With Subthreshold Leakage: Device/Circuit/Architecture Co-Design for Ultralow-Power Subthreshold Operation

Arijit Raychowdhury, *Student Member, IEEE*, Bipul C. Paul, *Senior Member, IEEE*, Swarup Bhunia, *Member, IEEE*, and Kaushik Roy, *Fellow, IEEE*

***Abstract*—This paper presents a novel design methodology for ultralow-power design using subthreshold leakage as the operating current (suitable for medium frequency of operation: tens to hundreds of millihertz). Standard design techniques suitable for superthreshold design can be used in the subthreshold region. However, in this study, it has been shown that a complete co-design at all levels of hierarchy (device, circuit, and architecture) is necessary to reduce the overall power consumption while achieving acceptable performance (hundreds of millihertz) in the subthreshold regime of operation. Simulation results of co-design on a five-tap finite-impulse-response filter shows $\sim 2.5\times$ improvement in throughput at iso-power compared to a conventional design.**

***Index Terms*—Parallelization and pipelining, pseudo-NMOS logic, subthreshold logic.**

I. Introduction

In RECENT years, the demand for power-sensitive designs has grown significantly. This tremendous demand has mainly been due to the fast growth of battery-operated portable applications such as notebook and laptop computers, personal digital assistants, cellular phones, and other portable communication devices. Further, due to the aggressive scaling of transistor sizes for high-performance applications, not only does subthreshold leakage current increase exponentially, but also gate leakage and reverse biased source–substrate and drain–substrate junction band-to-band tunneling (BTBT) currents increase significantly [1]–[3]. The unwanted tunneling currents may severely limit the functionality of the devices. In the VLSI system design space, considerable attention has been given to the design of medium/high-performance circuits (clock rates of several hundreds of millihertz) with power as a constraint. Well-known methods include: voltage scaling [4], [5], switching activity reduction [6], [7], architectural techniques such as pipelining and parallelism, and computer-aided design (CAD) issues of device sizing, interconnect [8], [9], and logic [10], [11] optimization.

Manuscript received December 13, 2004; revised June 6, 2005. This work was supported in part by the Semiconductor Research Corporation, Gigascale Silicon Research Centre, and the Defense Advanced Research Projects Agency.

A. Raychowdhury, B. C. Paul, and K. Roy are with the Department of Electrical and Computer Engineering, Purdue University, West Lafeyette, IN 47907 USA (e-mail: araycho@purdue.edu; paulb@purdue.edu; kaushik@purdue.edu).

S. Bhunia is with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: swarup.bhunia@case.edu).

Digital Object Identifier 10.1109/TVLSI.2005.859590

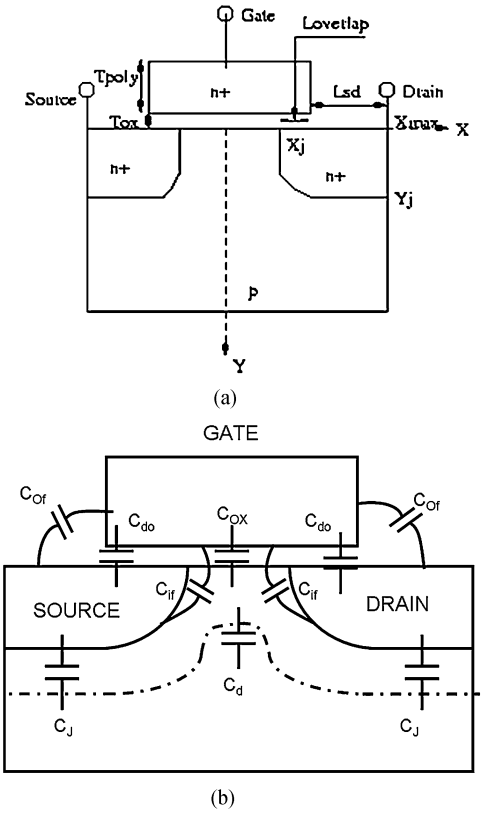
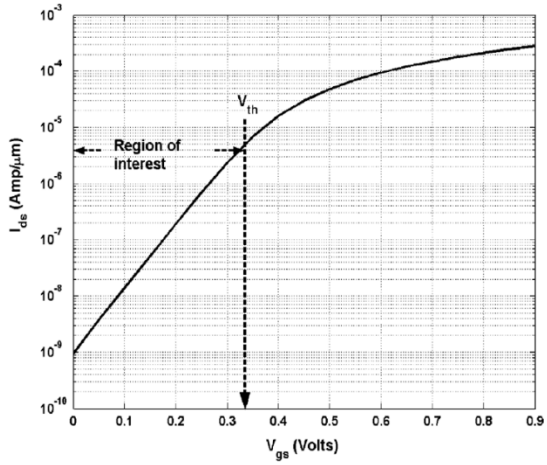


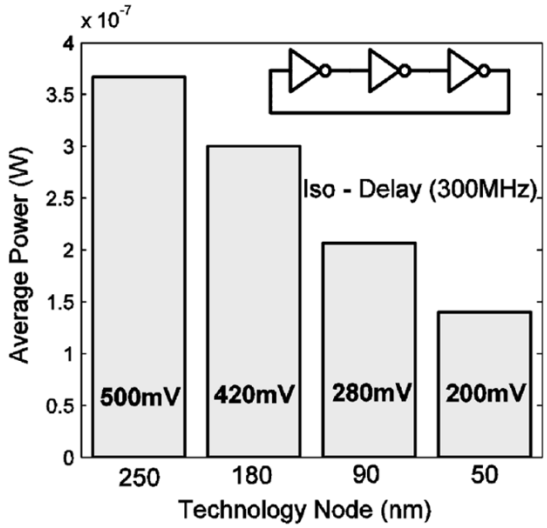
Fig. 1. (a) Architecture of the device. (b) Schematic showing the different capacitance components.

However, these methods are not sufficient in many applications such as portable computing gadgets medical electronics, where ultralow power consumption with medium frequency of operation (tens to hundreds of millihertz) is the primary requirement. To cope with this, several novel design techniques have been proposed. Energy recovery or quasi-adiabatic techniques promise to reduce power in computation by orders of magnitude. However, this involves the use of high-quality inductors, which makes integration difficult [12]. More recently, design of digital subthreshold logic was investigated with transistors operated in the subthreshold region (supply voltage $V_{\rm dd}$ corresponding to Logic 1, less than the threshold voltage $V_{\rm th}$ of the transistor) [13], [14]. In this technique, the subthreshold leakage current of the device is used for computation.

Due to the exponential I–V characteristics of the transistor [the schematic is shown in Fig. 1(a)], subthreshold logic gates



(a)



(b)

Fig. 2. (a) Region of operation of digital subthreshold logic. (b) Average power consumption of an inverter driving another inverter for different technology nodes. Note that all of the inverters are operating at iso-performance (a delay of 3.4 ns for each stage). The corresponding $V_{\rm dd}$'s used to achieve the performance are 500 mV (250 nm), 425 mV (180 nm), 280 mV (90 nm), and 200 mV (50 nm). The 250- and 180-nm devices are standard TSMC devices and the 90 and 50 nm are obtained from [15]. Note that, with technology scaling, the average power consumption decreases in the subthreshold regime at iso-transistor performance.

provide near ideal voltage transfer characteristics (VTCs). Furthermore, in the subthreshold region, the transistor input capacitance is less than that of strong inversion operation. The transistor input capacitance, C_i in subthreshold is a combination of intrinsic [oxide capacitance ($C_{\rm ox}$) and depletion capacitance (C_d)] and parasitic [overlap capacitance ($C_{\rm do}$), fringing capacitance ($C_{\rm if}, C_{\rm of}$)] [see Fig. 1(b)] and is given by [15]

$$C_i = \text{series}(C_{\rm ox}, C_d) \| C_{\rm if} \| C_{\rm of} \| C_{\rm do} \|. \quad (1)$$

In contrast, the input capacitance in strong inversion operation is dominated by the oxide capacitance. Due to the smaller capacitance and lower supply voltage (<threshold voltage of the transistor, $V_{\rm th}$), digital subthreshold circuits consume less power than their strong inversion counterpart at a particular frequency of operation. However, since the subthreshold leakage current is used as the operating current in subthreshold operation, these circuits cannot be operated at very high frequencies. Fig. 2(a) illustrates the region of operation for digital subthreshold operation.

Technology scaling in modern-day VLSI circuits has resulted in an exponential growth in the transistor performance at the sacrifice of both dynamic as well as static power. If performance is not the key, then the obvious doubt is whether scaling (vertical as well as lateral) is necessary for subthreshold logic. Since the target applications of subthreshold logic comprise of clock frequencies in the range of tens to hundreds of millihertz, it might appear that an aggressively scaled technology is unnecessary for its implementation. A 500- or 250-nm technology node can deliver the required performance for these applications. However, technology scaling reaps the high benefits of: 1) oxide thickness ($T_{\rm ox}$) scaling and 2) length scaling. As the oxide is thinned, the gate control on the channel increases, thereby improving the subthreshold slope of the device. In other words, a scaled device can deliver a required $I_{\rm on}$ (at iso-$I_{\rm off}$) at a lower supply voltage ($V_{\rm dd}$) compared to a nonscaled device. Further, the reduced device length results in a reduction in the total capacitance (both gate capacitance as well as junction capacitances). Consequently, the power to switch from one state (on/off) to another (off/on), which is called dynamic power, reduces. Fig. 2(b) shows how technology scaling can reduce the total average power consumption of an inverter (driving an identical inverter) in the subthreshold region of operation, all operating at the same frequency. As the technology is scaled, the $V_{\rm dd}$ can be reduced for iso-performance [500 mV (250 nm), 425 mV (180 nm), 280 mV (90 nm), and 200 mV (50 nm) for a delay of 3.4 ns]. This, coupled with the reduced capacitance, results in a sharp decrease in the power consumption of the inverter chain, as can be seen from Fig. 2(b). Hence, technology scaling results in reduction in the power consumption and is essential even in the subthreshold domain.

In this new paradigm of computation with leakage, unfortunately, conventional wisdom can deliver low-power systems but fails to provide the optimal or near-optimal solution. For subthreshold operation, the lowest power for a given throughput can be achieved only by a complete co-design in all the aspects of device, circuit, and architecture design.

Fig. 3 illustrates the power versus throughput of a five-tap conventional finite-impulse response (FIR) filter (a typical signal processing application). On one hand, conventionally we have the high-performance and high-power regime. As the $V_{\rm dd}$ is scaled, both the throughput and power reduces until we reach the subthreshold domain of operation. In this paper, we will endeavor to improve the throughput of this system at iso-power in the subthreshold region by device/circuit/architectural techniques. To the best of our knowledge, for the first time, this paper investigates the modeling and optimization of digital subthreshold logic units from the aspects of devices, circuits, as well as architecture to provide ultralow-power digital operation suitable for the medium frequency range (tens to hundred of millihertz) of applications, Thus, we provide a complete methodology to digital subthreshold design to achieve minimum power consumption. We have selected the 50-nm technology to demonstrate our design methodology (by 50 nm, we mean the effective length $L_{\rm eff}$ of the device). This corresponds to a 90-nm technology node as defined by ITRS.

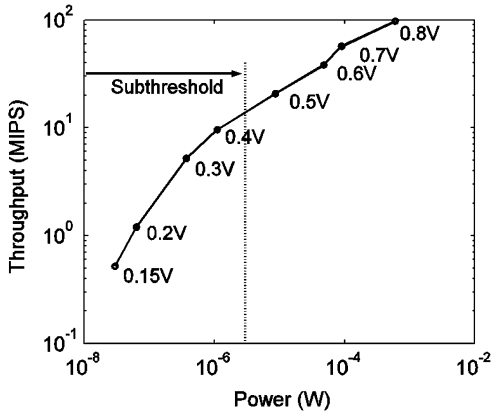


Fig. 3. Power-throughput tradeoff of a five-tap FIR filter. The corresponding $V_{\rm dd}$'s are shown.

The remainder of the paper is organized as follows. In Section II, we develop a device optimization methodology suited for subthreshold operation. The choice of circuit styles in the subthreshold domain is elaborated on in Section III. The architectural techniques to achieve higher throughput at lower power in the subthreshold region is discussed in Section IV. Finally, we bring forth the effects of variation in Section V and draw conclusions in Section VI.

II. Device Design Considerations

In all of the previous work, standard transistors were operated in the subthreshold region to implement subthreshold logic. Standard transistors are the "super-threshold transistors" that are used for ultrahigh-performance design. It is only prudent to investigate if the standard transistors are well suited for subthreshold operation. In [16], the authors have reported a device optimization strategy for subthreshold operation. This will be described here, in brief, for the convenience of the readers.

Scaled device technologies demand nonuniform doping profiles to provide good control on the electrical characteristics of the device. It is an established fact that, for scaled super-threshold transistors, it is essential to have halo and retrograde doping to suppress the short-channel effects. The main functions of halo doping and retrograde wells are to reduce drain-induced barrier lowering (DIBL), to prevent body punch-through, and to control the threshold voltage of the device independent of its subthreshold slope. However, in subthreshold operation, it is worthwhile to note that the overall supply bias is small (on the order of 150–400 mV). Consequently, the effects of DIBL and body punch-through are extremely low. Further, as long as we have a fixed $I_{\rm off}$, better subthreshold slope S implies a better device. Since our interest is in the region below the threshold voltage, the threshold voltage of the device is not of any interest to us as long as we meet a predefined $I_{\rm off}$ and S. Hence, it can be qualitatively argued that the halo and retrograde doping are not essential for subthreshold device design [3]. The absence of the halo and retrograde doping has the following implications:

- a simplified process technology in terms of process steps and cost;
- a significant reduction of the junction capacitances.

The halo regions near the source–substrate and the drain–substrate regions significantly increase the junction capacitances, thereby increasing the switching power and the delay of the logic gates. The absence of the halo/retrograde doping will reduce this junction capacitance.

It should, however, be noted that the doping profile in the optimized devices should have a high-to-low doping profile. It is necessary to have a low doping level in the bulk of the device to:

- reduced the capacitance of the bottom junction;
- reduced substrate noise effects and parasitic latch-up problems.

In order to evaluate the feasibility of possible device-level optimization, we start with the super-halo standard devices,[1] model subthreshold current, and junction capacitance and then optimize the doping profile for subthreshold operation. The device under consideration has an effective channel length of 50 nm and an oxide thickness of 2 nm.

A. Modeling Subthreshold Current ($I_{\rm ds}$) and Junction Capacitance (C_J)

In this section, we present a general approach to formulate the model for the junction capacitance, the depletion capacitance, and the subthreshold current of a MOSFET. The models will be used to evaluate the subthreshold circuits. The formulation, which was developed for NMOS transistors, can be easily extended to PMOS transistors. Device structures with Gaussian-shaped channel ("super halo" channel doping) and source/drain (S/D) doping profiles have been considered while deriving these models. A schematic of the device structure (symmetric about the middle of the channel) is shown in Fig. 1(a). The two-dimensional (2-D) Gaussian doping profile in the channel $(N_a(x,y))$ and S/D $(N_{\rm sd}(x,y))$ can be represented as [17]

$$\begin{aligned} &x>0\\ &N_a(x,y)=A_p\Gamma_{xa}(x)\mathrm{K}_{ya}(y)+N_{\rm SUB}\\ \text{where}\quad &\mathrm{K}_{ya}(y)=\exp\left(\frac{-(y-\alpha_a)^2}{\sigma 1_{ya}^2}\right), \qquad 0\le y\le\alpha_a\\ &\qquad\quad =\exp\left(\frac{-(y-\alpha_a)^2}{\sigma 2_{ya}^2}\right), \qquad y>\alpha_a \end{aligned} \tag{2}$$

and

$$\left[\begin{array}{ll}\Gamma_{xa}(x) &=\exp\left(\frac{-(x-\beta_a)^2}{\sigma_{xa}^2}\right), \quad 0\le x\le\beta_a\\ &=1, \qquad x>\beta_a\end{array}\right].$$

Similarly, the S/D doping $(N_{\rm sd}(x,y))$ can be represented as

$$\begin{aligned} &x>0\\ &N_{\rm sd}(x,y)=A_{\rm sd}\Gamma_{x\,\rm sd}(x)\mathrm{K}_{y\,\rm sd}(y)\end{aligned}$$

where

$$\mathrm{K}_{y\,\rm sd}(y)=\exp\left(\frac{-(y-\alpha_a)^2}{\sigma 1_{y\,\rm sd}^2}\right) \tag{3}$$

and

$$\left[\begin{array}{ll}\Gamma_{x\,\rm sd}(x) &=\exp\left(\frac{-(x-\beta_{\rm sd})^2}{\sigma_{x\,\rm sd}^2}\right), \qquad 0\le x\le\beta_{\rm sd}\\ &=1, \qquad x>\beta_{\rm sd}\end{array}\right]$$

[1][Online]. Available: http://www-mtl.mit.edu/Well

where $A_{\rm p}$ and $A_{\rm sd}$ represent the peak "halo" and S/D doping, respectively, and $N_{\rm SUB}$ is the constant uniform doping in the bulk and is much less compared to contributions from Gaussian profiles at and near the channel and S/D regions. Parameters $\sigma 1_{ya}, \sigma 2_{ya}, \alpha_a, \alpha_a \alpha_{\rm sd} (=0), \beta_a$, and $\beta_{\rm sd}$ control the positions and $\sigma 1_{ya}, \sigma 2_{ya}, \sigma_{xa}$, and $\sigma_{y\,{\rm sd}}, \sigma_{x\,{\rm sd}}$ control the variances of the Gaussian profiles in channel and S/D regions [17]. Unless otherwise specified in this paper, we have used NMOS $({\rm N_{ref}})$ and PMOS $({\rm P_{ref}})$ transistors with $L_{\rm eff} = 50$ nm, $W_{\rm eff} = 1\ \mu$m, and channel doping profile $\alpha_a = 0.018\ \mu$m, $\sigma 1_{ya} = 0.018\ \mu$m $\sigma 2_{ya} = 0.018\ \mu$m, $\beta_a = 0.016\ \mu$m, $\sigma_{xa} = 0.016\ \mu$m, and the S/D profile from footnote 1.

In the subthreshold state of a device $(V_{\rm gs} < V_{\rm th})$, the current flowing from the drain to the source of a transistor is known as the subthreshold current. The subthreshold current flowing through a transistor is given by [15]

$$I_{\rm sub} = \frac{W_{\rm eff}}{L_{\rm eff}} \mu \sqrt{\frac{q\varepsilon_{si} N_{\rm cheff}}{2\Phi_s}} \nu_T^2 \times \exp\left(\frac{V_{\rm gs} - V_{\rm th}}{m\nu_T}\right)\left(1 - \exp\left(\frac{-V_{\rm ds}}{\nu_T}\right)\right) \quad (4)$$

where ${\rm N_{cheff}}$ is the effective channel doping, s is the surface potential, m is the body-effect coefficient related to the subthreshold swing, and v_T is the thermal voltage given by kT/q. Using the charge sharing model and following the procedure given in footnote 1, the threshold voltage can be expressed as

$$V_{\rm th} = V_{\rm FB} + (\Phi_{s0} - \Delta\Phi_s) + \gamma\sqrt{\Phi_{s0} - V_{\rm bs}}\left(1 - \lambda\frac{X_d}{L_{\rm eff}}\right) + \Delta V_{\rm NWE} \quad (5)$$

where $V_{\rm FB}$ is the flat-band voltage, Φ_{s0} is the zero bias surface potential, γ is the body factor, $C_{\rm ox} = \varepsilon_{\rm sio2}/{\rm tox}$ is the oxide capacitance, $X_{\rm d}$ is the depletion layer thickness, λ is a fitting parameter (~ 1), and $\Delta V_{\rm NWE}$ is the narrow-width correction factor given in [15]. $\Delta\Phi_{\rm s}$ is the reduction of the surface potential (Φ_s) of short-channel devices from its zero bias value due to short-channel effects like DIBL and $V_{\rm th}$ roll-off [15]. Different parameters in the above model depend on the effective channel and S/D doping [15]. We have evaluated the effective channel $(N_{\rm cheff})$ and S/D doping $(N_{\rm sdeff})$ considering the exact 2-D Gaussian doping profile [given in (2), (3)] given as follows:

$$\begin{aligned} N_{\rm sdeff} &= \frac{1}{\Delta_{\rm sd}} \int_{\Delta} \int_{\rm sd} N_{\rm sd}(x,y)\,dx\,dy \\ &= \frac{A_{\rm sd}}{\Delta_{\rm sd}} \int_{x=-X_j}^{x=L_{\rm gate}/2+L_{\rm sd}} \Gamma_{x\,{\rm sd}}(x)\,dx \int_{y=0}^{y=y_j} {\rm K}_{y\,{\rm sd}}(y)\,dy \\ N_{\rm cheff} &= \frac{1}{\Delta_{\rm ch}} \int_{\Delta} \int_{\rm ch} N_a(x,y)\,dx\,dy + N_{\rm sub} \\ &= \frac{A_p}{\Delta_{\rm ch}} \int_{x=-L_{\rm eff}/2}^{x=+L_{\rm eff}/2} \Gamma_{xa}(x)\,dx \int_{y=0}^{y=X_d} {\rm K}_{ya}(y)\,dy + N_{\rm sub} \end{aligned} \quad (6)$$

where $\Delta_{\rm SD} = (L_{\rm overlap} + L_{\rm sd})Y_j$ is the S/D area, $L_{\rm overlap}$ is the gate and the S/D overlap length, and $L_{\rm sd}$ is the S/D length as shown in Fig. 1(a).

$\Delta_{\rm ch} = L_{\rm eff} X_d$ is the area of the channel region which is under the influence of gate. To calculate the effective doping, X_d is assumed to be α_a since most of the depletion charge is confined in the region $y = 0$ to $y = \alpha_a$. Since we have considered the exact Gaussian nature of the doping profile (instead of approximating it as step profile as described in [15]), we have been able to capture the effect of change in the doping profile more accurately.

The inverse subthreshold slope (S) for the short-channel device, considering the penetration of the drain-induced electric fields in the center of the channel is given by

$$S \approx 2.3\frac{mkT}{q}\left(1 + \frac{11 t_{\rm ox}}{X_d} e^{-\pi L/(x_d + 3t_{\rm ox})}\right) \quad (7a)$$

where the body effect coefficient m is given by

$$m = 1 + \frac{3t_{\rm ox}}{X_d}. \quad (7b)$$

For a symmetric device, the capacitance expressions for the drain and source junctions will be identical. Hence, here we have considered only the drain junction. The junction capacitance is given by

$$C_j = \frac{w_{\rm eff} l \varepsilon_{\rm si}}{W_{dj}} = w_{\rm eff} l \sqrt{\frac{\varepsilon_{\rm si} q}{2(\psi_{\rm bi} + V_j)}\left(\frac{N_a N_d}{N_a + N_d}\right)} \quad (8)$$

where $N_{\rm a}$ and $N_{\rm d}$ are the doping concentrations on the two sides of the junction, $\psi_{\rm bi}$ is the built-in potential across the junction, and $V_{\rm j}$ is the potential applied across the junction. The length l is the length of the junction and $w_{\rm eff}$ is the effective width of the NMOS. To obtain an accurate analytical estimate of the total junction capacitance, we can apply the "rectangular junction" approximation. Using this approximation, the total capacitance at the drain junction is given by

$$\begin{aligned} C_{J-{\rm drain}} &= C_{\rm side} + C_{\rm bottom} \\ &= w_{\rm eff}(x_2 - x_1)\sqrt{\frac{\varepsilon_{si} q}{2(\psi_{bi} + V_j)}\left(\frac{N_{a{\rm side}} N_{d{\rm side}}}{N_{a{\rm side}} + N_{d{\rm side}}}\right)} \\ &\quad + w_{\rm eff}(y_2 - y_1)\sqrt{\frac{\varepsilon_{si} q}{2(\psi_{bi} + V_j}\left(\frac{N_{a{\rm bottom}} N_{d{\rm bottom}}}{N_{a{\rm bottom}} + N_{d{\rm bottom}}}\right)} \end{aligned} \quad (9)$$

where $N_{\rm aside}$ and $N_{\rm dside}$ are given by

$$\begin{aligned} N_{a{\rm side}} &= \frac{1}{|y_2 - y_1|}\int_{y1}^{y2} N_a(X_j, y)\,dy \\ N_{d{\rm side}} &= \frac{1}{|y_2 - y_1|}\int_{y1}^{y2} [N_{\rm sd}(x = \beta_{\rm sd}, y) - N_a(x = \beta_a, y)]\,dy. \end{aligned} \quad (10)$$

$X_{\rm j}$ and $Y_{\rm j}$ are found by solving the following equations:

$$\begin{aligned} N_{\rm sd}(X_j, y = 0) &= N_a(X_j, y = 0) \\ N_{\rm sd}(x = x_{\max}, y_j) &= N_a(x_{\max}, Y_j). \end{aligned} \quad (11)$$

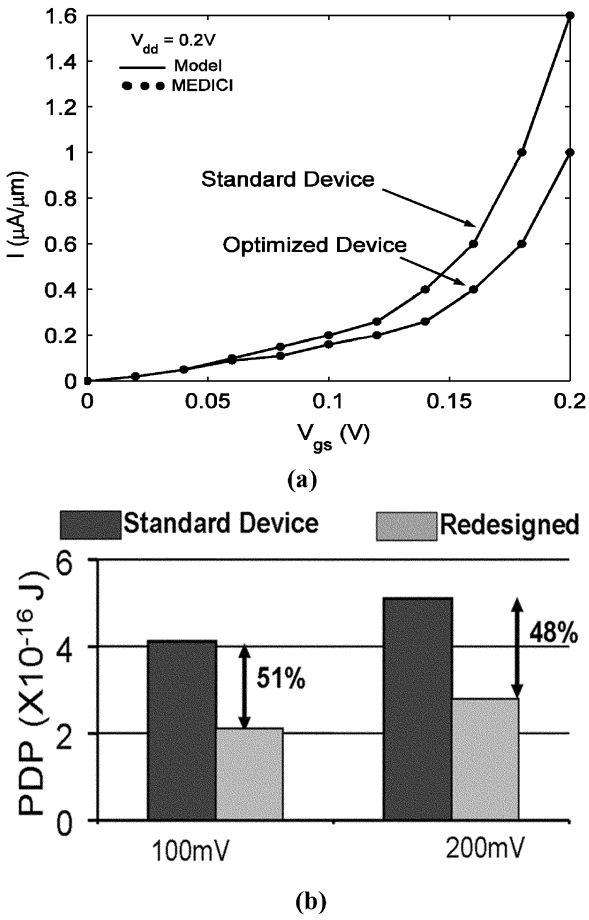


Fig. 4. (a) ID—$V_{\rm GS}$ characteristics of the optimized and the standard device. (b) Power–delay product (PDP) of an inverter (driving an identical inverter) with standard and redesigned optimized devices.

For simplicity, we have considered $y1 = 0$ and $y2 = Yj$ (for the side junction) and $x1 = X_j$ and $x2 = x_{\max}$ (for the bottom junction).

B. Device Optimization

It has already been qualitatively argued that, for typical subthreshold operation, the highly doped halo regions can be removed. This would increase the $I_{\rm off}$ and for very low values of the overall doping concentration the subthreshold slope too will be affected. The increase in $I_{\rm off}$ and S can be compensated for by increasing $\sigma 1{\rm ya}$, and, for the simplified process technology, it will be useful if the high-to-low doping profile ($\sigma_{1{\rm ya}} = \infty$) can be used. For the purpose of our comparison, we keep the $I_{\rm off}$ constant. A channel doping concentration of 3.8×1018 cm^{-3} gives an OFF current of 1 nA/μm at $V_{\rm dd} = 200$ mV (the same as the standard device). Further, the subthreshold slope of this optimized device ($\sim$83 mV/decade) is better than that of the standard device ($\sim$90 mV/decade). This can be noted by comparing the currents of the two devices in Fig. 4(a). Another important aspect for the optimized subthreshold device is the junction capacitance. The junction capacitance of the optimized device is lower than the standard device because of lower overall junction doping and the removal of halo doping. In comparison to $C_J = 4.9 \times 10$–16 F/μm for the standard device, the junction capacitance (by analytical model and verified using MEDICI simulations) is $C_J = 3.2 \times 10$–16 F/μm for the optimized device. Hence, the junction capacitance decreases by approximately 35%. The intrinsic device capacitance [oxide capacitance $(C_{\rm ox})$ and depletion capacitance $(C_{\rm d})$], however, remains unaffected. The depletion capacitance remains the same because the depletion region extends until αa, which has not been changed here. This results in a reduction of $\sim$17% of the total capacitance as defined in (1). Thus, the optimized subthreshold device has a better S and a lower junction capacitance at iso-$I_{\rm off}$ conditions.

Note that this optimized device is suited for subthreshold operation only. In the super-threshold region, these devices will not work because of extremely high SCE (due to the absence of halo implants) at higher $\rm V_{dd}$. In the optimized subthreshold device, by removing the super halo regions (since lower $\rm V_{dd}$ reduces DIBL and SCE), we have effectively reduced the junction capacitances and improved the subthreshold slope at iso-$I_{\rm off}$. Consequently, we get significant improvement in the PDP of circuits built with the optimized devices as compared to the standard devices. Fig. 4(b) illustrates how the PDP of an inverter driving an identical inverter can be reduced by using the optimized devices. By PDP, we mean the switching energy, i.e., the energy required to switch the state from ZERO to ONE or vice versa.

III. Circuit Design Considerations

In this section, we will seek further improvement in PDP by proper choice of circuit style in subthreshold domain. We will first establish the superiority of pseudo-NMOS logic over conventional CMOS logic in the subthreshold region (Section III-A) and propose analytical models for computing gate-level delay and power (Sections III-B and III-C).

A. Subpseudo-NMOS Logic

Pseudo-NMOS logic is faster than static CMOS due to smaller load capacitance and shorter interconnects. However, in order to utilize pseudo-NMOS logic, the drawbacks of ratioed logic such as large static current consumption and degradation in static noise margin should be carefully taken into account. Pseudo-NMOS logic in the subthreshold region (subpseudo-NMOS) inherits the advantages it has in the strong inversion such as higher performance and smaller area. In addition to this, the drawbacks of ratioed logic are relieved in the subthreshold region [18]. This is mainly because, in the subthreshold region, the drain current saturates and becomes independent of $V_{\rm ds}$ for $V_{\rm ds} > \sim 3kT/q$ (78 mV at 300 K). Note that a transistor in strong inversion region (super-threshold) only enters the saturation region when $V_{\rm ds} > V_{\rm gs} - V_{\rm th}$, which gives a much narrower saturation region and, thus, an undesirable voltage transfer characteristic (VTC). However, the output voltage of pseudo-NMOS in subthreshold region swings for almost rail-to-rail and thus provides a high noise margin. This can be noted from Fig. 5. When both the super-threshold and subthreshold pseudo-NMOS gates are sized for identical low-to-high and high-to-low delays, the subthreshold gate provides a better noise margin and a sharper VTC.

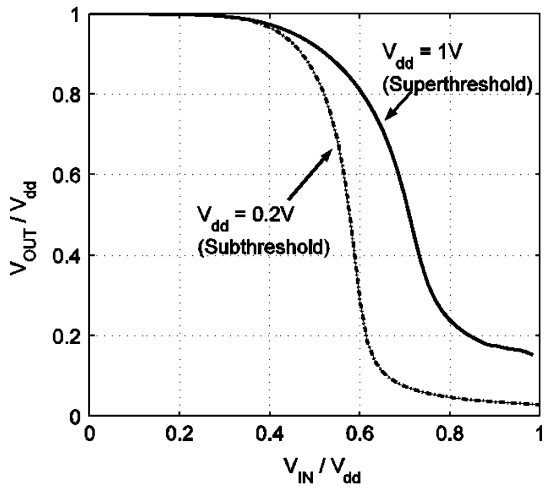


Fig. 5. VTC of a subthreshold and a super-threshold transistor designed using pseudo-NMOS logic. Both gates have been sized for minimum delay. Note that the subthreshold transistor has a higher gain and a better noise margin than the super-threshold design.

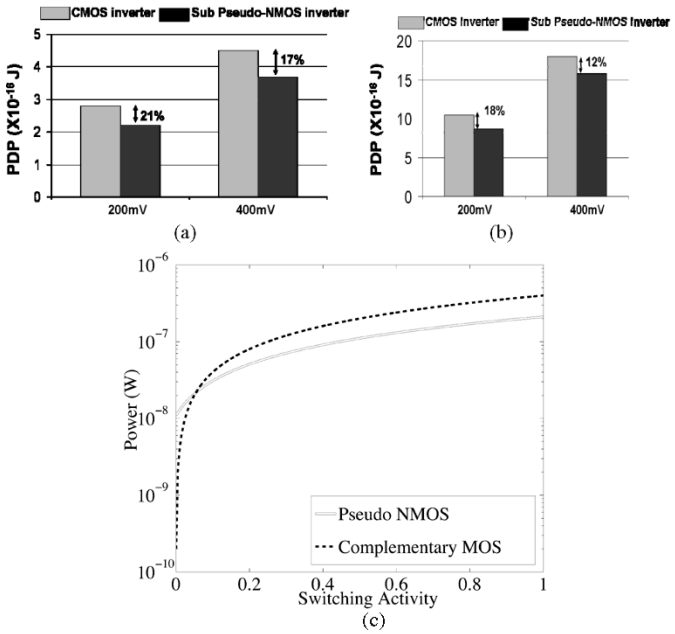


Fig. 6. PDP of an inverter in CMOS and subpseudo-NMOS logic styles made with the optimized devices driving (a) an identical inverter, (b) four identical inverters, and (c) the total power of an inverter driving an identical inverter as a function of switching activity.

Subpseudo-NMOS is also more efficient than sub-CMOS in terms of PDP. Simulation results of a pseudo-NMOS inverter (driving an identical inverter as well as a fan-out of four) and a CMOS inverter are compared in Fig. 6(a) and (b). We observe that, in the subthreshold region, pseudo-NMOS gives approximately 20% improvement in PDP compared to CMOS. The reason behind the lower PDP in subpseudo-NMOS is the smaller delay. Further, in the subthreshold region, the static short circuit current is also a weak inversion current, which is relatively much less significant than in super-threshold. Therefore, we propose the use of pseudo-NMOS logic in the subthreshold domain.

However, it should be noted that the pseudo-NMOS logic style suffers from a static leakage when the pull-down network

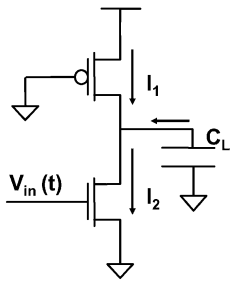


Fig. 7. Different current components in the inverter when $V_{\rm in}(t)$ ramps up.

in ON. In the subthreshold domain, however, this ON current is orders of magnitude smaller than super-threshold. Hence, in cases where the switching activity is extremely low, the total power consumption in the pseudo-NMOS logic gate can be potentially higher than CMOS. Fig. 6(c) plots the total power consumption of an inverter driving an identical inverter for varying switching activities. It can be noted that, for switching activity above 6%, pseudo-NMOS logic operates at lower power consumption than CMOS.

B. Modeling Delay of Primitive Logic Gates

In order to evaluate the implications of the device-level optimization on circuit and architecture, it is necessary to model the delay and power (both dynamic as well as static) of the primitive logic gates. It should be noted that, in the subthreshold region, the leakage is the drain current at $V_{\rm GS}$=0 and the ON current is at $V_{\rm GS} = V_{\rm DD}(<V_{\rm th})$. Let us consider an inverter (shown in Fig. 7) with an output load capacitance of CL. Let us first assume that CL is constant. We have previously argued that, for subthreshold operation, the device I–V parameters of interest are $I_{\rm off}$ and S. $I_{\rm off}$ $(@V_{\rm ds} = V_{\rm dd})$ for the optimized NMOS device is 1 nA/μm and $S \sim 83$ mV/decade at $V_{\rm dd} = 200$ mV. For the corresponding PMOS device, we have engineered the gate work function to obtain the same $I_{\rm off}$ $(@V_{\rm ds} = V_{\rm dd})$ as NMOS. Now, consider a ramp input $V_{\rm in}(t)$ (going from 0 to $V_{\rm dd}$) to the NMOS of a pseudo-NMOS inverter. Let us denote the subthreshold current which is a function of $V_{\rm ds}$ and $V_{\rm gs}$ (4) by $I = {\rm Isub}(V_{\rm ds}, V_{\rm gs})$. The discharging current $(I2 - I1)$ of the load capacitor C_L is obtained by self -consistently solving the following equations:

$$I_1 = I_{\rm sub}^{\rm PMOS}(V_{\rm dd}, V_{\rm out} - V_{\rm dd})$$
$$I_2 = I_{\rm sub}^{\rm NMOS}(V_{\rm in}(t), V_{\rm out}) \tag{12a}$$
$$I_1 - I_2 = C_L \frac{dV_{\rm OUT}}{dt}. \tag{12b}$$

The delay (T_d) of the inverter can be obtained from the following:

$$\int_0^{T_d} (I_1 - I_2)\, dt = \int_{V_{\rm out}=V_{\rm dd}}^{V_{\rm out}=V_{\rm dd}/2} C_L\, dV_{\rm OUT}. \tag{13}$$

The same technique can be used to evaluate the low-to-high transition at the inverter output. Note that, in reality, C_L is not a constant but actually depends on $V_{\rm OUT}$. It is evaluated by simply adding the voltage-dependent input capacitance of the load C_i and the junction capacitances of the driver transistors

$$C_L = C_J^{\rm NMOS} + C_J^{\rm PMOS} + C_i \tag{14}$$

where C_i is given by (1). For multiple fan-out gates, the C_i is suitably adjusted.

The same methodology can be used for evaluating the delay of NAND and NOR gates. In case of a NAND gate, we assume I2 to be an equivalent discharging current through the NMOS stack. The role of charge sharing by the internal node in case of the NAND stack is incorporated by assuming

$$C_L = \beta^* \left(C_J^{\mathrm{NMOS}} + C_J^{\mathrm{PMOS}}\right) + C_i \tag{15}$$

where β is a capacitance fitting parameter.

Fig. 8 illustrates how the delay of the three primitive gates depends on the V_{dd}. It can be noted that the simulation results from the model described above closely match with MEDICI simulation. It can be noted that the subpseudo-NMOS gates (with the optimized devices) have minimum delay.

C. Modeling Power of Logic Gates

The total power of the logic gate consists of a dynamic component and a static component. In our model, we have used

$$\begin{aligned} P_{\mathrm{Total}} &= P_{\mathrm{Dynamic}} + P_{\mathrm{Static}} \\ P_{\mathrm{Total}} &= \alpha C_L V_{\mathrm{dd}}^2 f + I_{\mathrm{Leak}} V_{\mathrm{dd}} \end{aligned} \tag{16}$$

where α is the activity of the logic gate, f is the clock frequency, and I_{Leak} is its leakage current.

IV. Architectural Considerations

In the previous sections, we have deduced that, for subthreshold operation, it is prudent to optimize the transistors so as to reduce the junction capacitance (and make the subthreshold slope better). From a circuit perspective, we have demonstrated that subpseudo-NMOS has an advantage in terms of PDP over conventional CMOS. In this section, we explore the design space at the architectural level. Since the primary objective of system design in subthreshold operation is low power, we focus on searching the design space for minimizing total power while maintaining a target throughput (or maximizing throughput at a constant power). Further, it may not be possible to meet a target throughput by device and circuit optimization only. Hence, it is necessary to exploit architectural level techniques to obtain higher throughput at lower power.

Parallelism and pipelining are two important ways to improve system performance during architectural synthesis. Typically, they represent a tradeoff between area/power and system throughput. Parallelism is associated with incorporating more hardware resources to increase throughput. However, more parallel elements result in increased leakage power and die area. On the other hand, the idea of a pipelined design is associated with increasing operating frequency (and hence throughput), by breaking the critical signal propagation path of a circuit into multiple smaller paths (with latch insertion at appropriate places) [19]. However, the increase in number of pipeline stages corresponds to an increase in overhead due to latches (in terms of die area and power) and latency of operation.

The reduction in supply voltage is a popular choice to reduce power dissipation since it renders quadratic savings in dynamic power. The performance loss due to supply voltage scaling can be compensated by architectural means using more parallel resources or increasing the number of pipeline stages as mentioned above. However, in the subthreshold region, the dynamic power constitutes a smaller percentage of the total power than in super-threshold. This can be understood from the following:

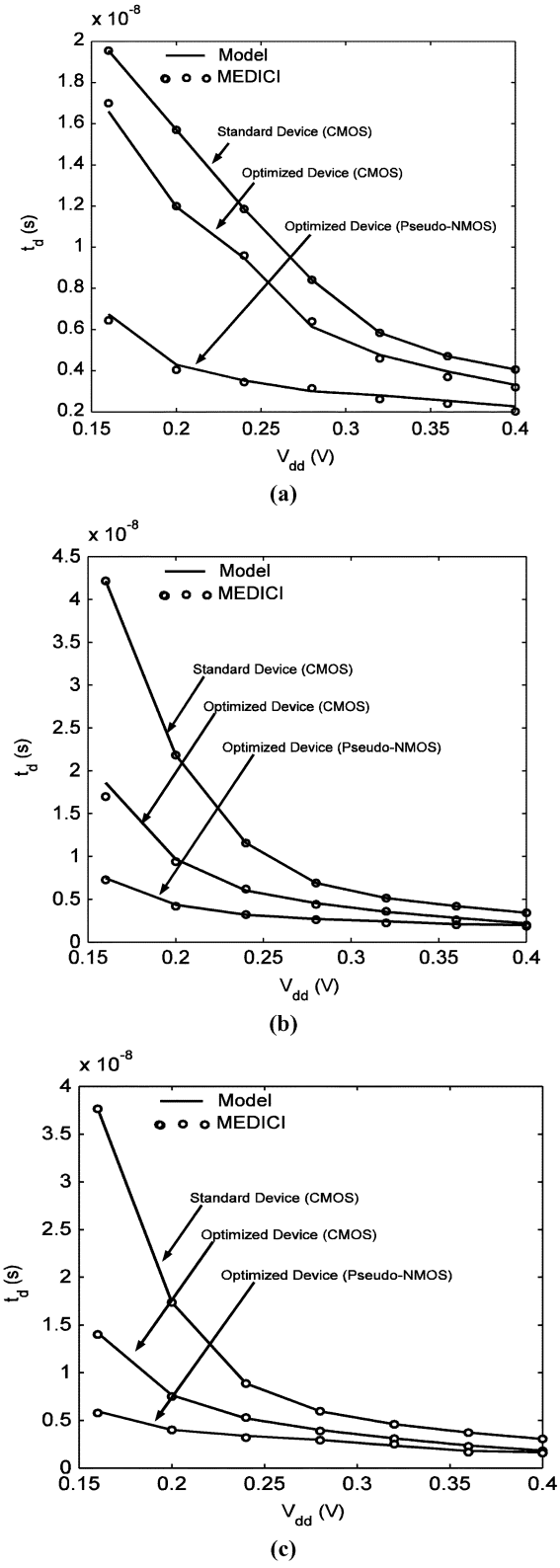


Fig. 8. Delay versus V_{dd} of (a) an inverter, (b) a NAND gate, and (c) a NOR gate. The load for all of the gates is an inverter.

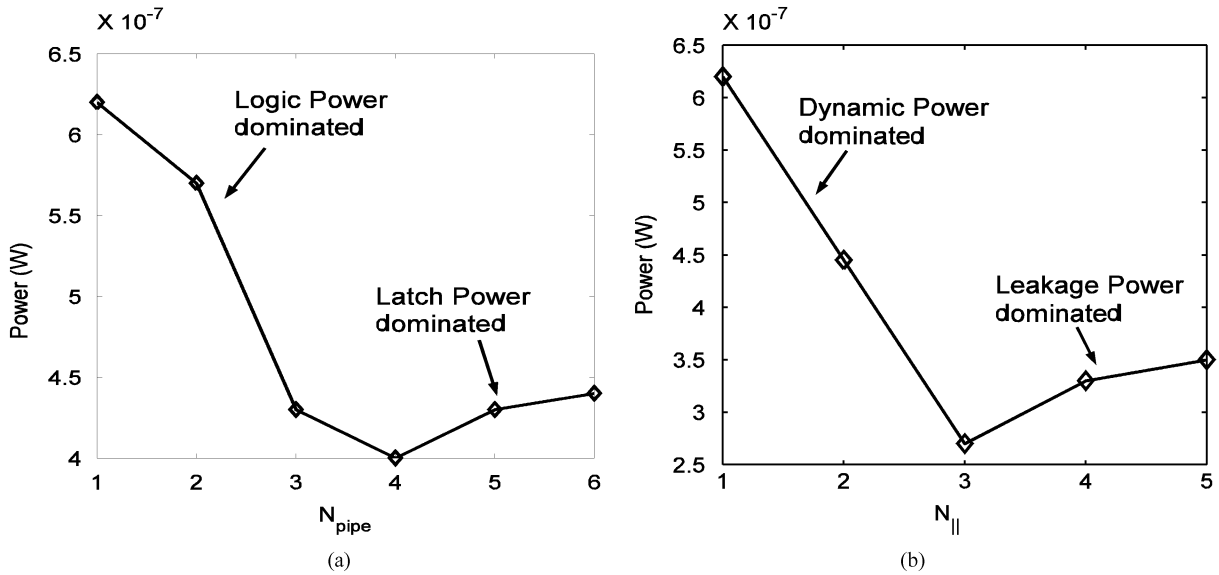


Fig. 9. For a five-tap FIR filter at 10 MIPS: (a) variation of power with number of pipeline stages (one parallel unit) and (b) variation of power with number of parallel units (one pipeline stage). $V_{\rm dd}$ has been adjusted in each case to achieve the right throughput.

- the gate input capacitance is lower than its super-threshold counterpart;
- the operating voltage $V_{\rm dd}$ as well as the operating frequency f are lower than in the super-threshold design.

Even for the same frequency of operation, a subthreshold design has lower dynamic power than the super-threshold design. Consequently, even in the active mode of the circuit, the leakage power is a significant portion of the total power.

To maintain a constant throughput, we can lower $V_{\rm dd}$ if we incorporate more parallel processing units. This, of course, decreases the dynamic power of the system. However, the leakage power increases steadily (because of more number of processing units) and very soon becomes a dominant portion of the total power. In subthreshold operation this trade-off becomes apparent for a relatively lesser number of parallel units. Hence, contrary to belief that parallelization can ideally reduce power for constant throughput (by aggressively scaling $V_{\rm dd}$), in subthreshold it is necessary to judiciously choose the number of parallel units such that the total power $(\mathrm{dynamic} + \mathrm{leakage})$ is minimized. This can be noted from Fig. 9(a).

Pipelining also offers a power–frequency tradeoff in terms of the contributions from the latches. The latches contribute significantly not only to the dynamic but also to the leakage power. Hence, the number of pipeline stages need to be chosen so that the total power $(\mathrm{combinational\ logic} + \mathrm{latches})$ is minimized [as shown in Fig. 9(b)]. Thus, there is a need for identifying the global optimal in terms of number of parallel units and pipeline stages as depicted in Fig. 10. The total system power can be estimated as

$$P_{\rm Total}^{\rm System} = N_{||}\left[\begin{array}{l}\left(\alpha^{\rm comb}C_{\rm total}^{\rm comb}V_{\rm dd}^2 f + I_{\rm Leak}^{\rm comb}V_{\rm dd}\right) \\ \quad + K^* N_{\rm stage}\left(\alpha^{\rm Latch}C_{\rm total}^{\rm Latch}V_{\rm dd}^2 f + I_{\rm Leak}^{\rm Latch}V_{\rm dd}\right)\end{array}\right] \tag{17}$$

where $N_{||}$ is the number of parallel units, $N_{\rm stage}$ is the number of pipeline stages, and K is the number of latches required per stage.

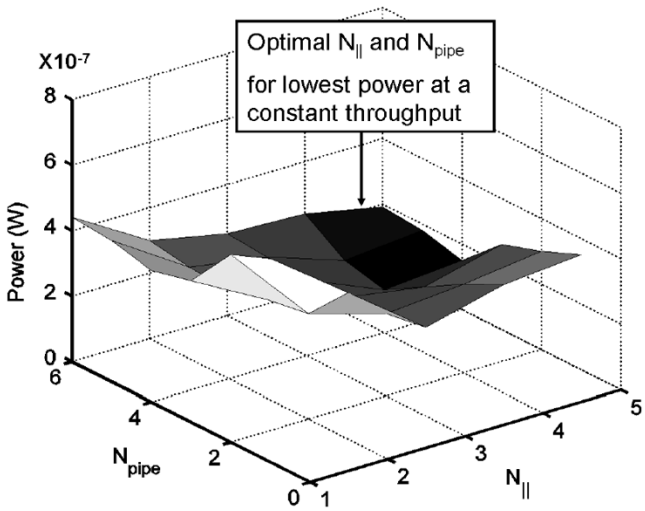


Fig. 10. Choice of optimal $N_{||}$ and $N_{\rm pipe}$ for lowest power consumption at a target throughput (= 10 MIPS in this case for a five-tap FIR filter). Note the global minima.

Fig. 11 illustrates the flow-diagram for choosing the optimal number of $N_{||}$ and $N_{\rm stage}$ so that the total system power for a target throughput is minimized. Let us consider the 8-b five-tap FIR filter whose power-performance tradeoff was previously illustrated in Fig. 3. We have used the optimized devices to implement subpseudo-NMOS logic. Further, the design has been optimally pipelined and parallelized. This entails considerable improvement in throughput for iso-power, as illustrated in Fig. 12. First, at iso-power, we improve the throughput by device optimization and then by switching to pseudo-NMOS. Next, we implemented an optimal pipelining/parallelization strategy to increase throughput. It can be noted that the throughput obtained in the process is two to nine times higher that of the conventional design (at iso-power). We have also applied the optimal parallelization/pipelining techniques to the conventional FIR design in the subthreshold region to achieve

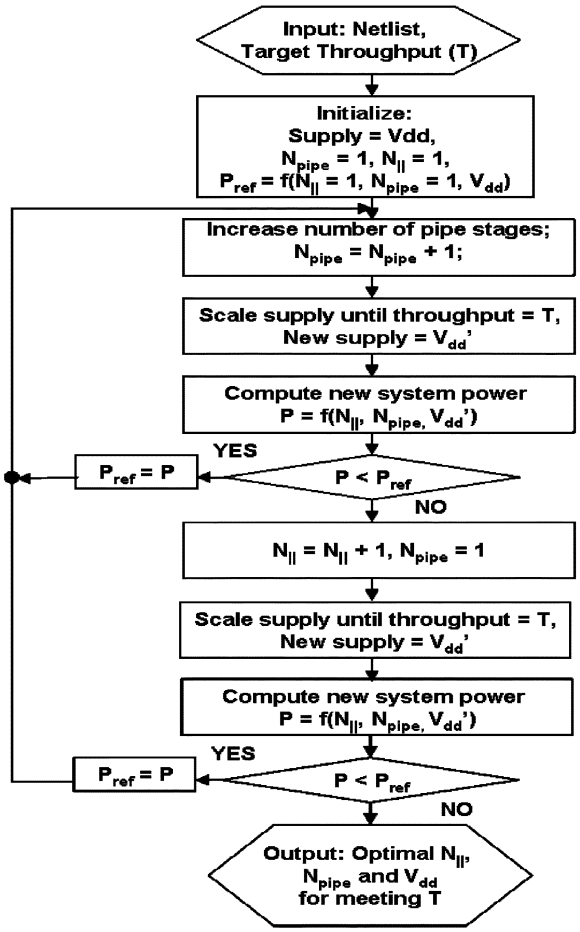


Fig. 11. Flow diagram for choosing the optimal number of parallel units ($N_{||}$) and pipeline stages (N_{stage}) for minimum power at a target throughput.

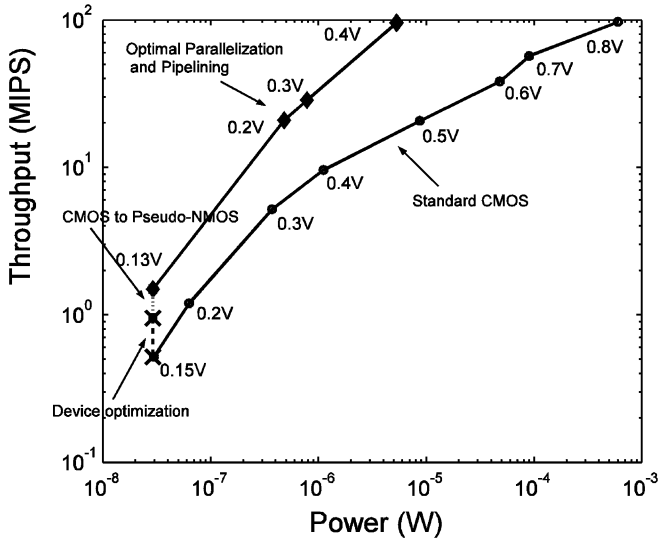


Fig. 12. Throughput versus power of a five-tap FIR filter (refer to Fig. 1). Note how device optimization, choice of circuit style, and optimal parallelization/pipelining can significantly improve throughput at iso-power.

higher throughput. Nevertheless, with the device/circuit/architectural optimizations discussed so far, the throughput obtained

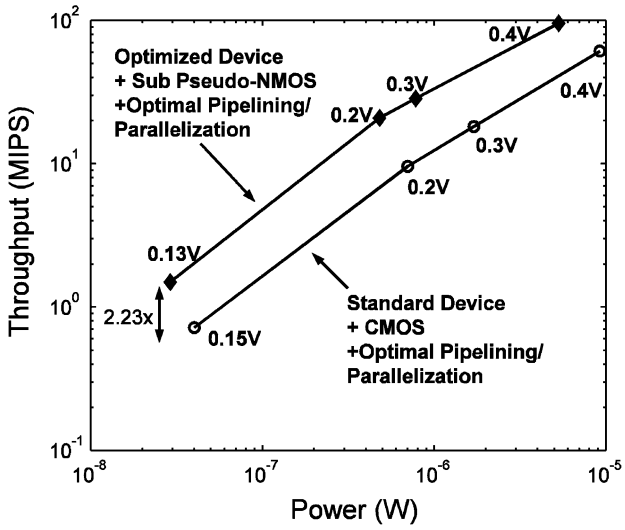
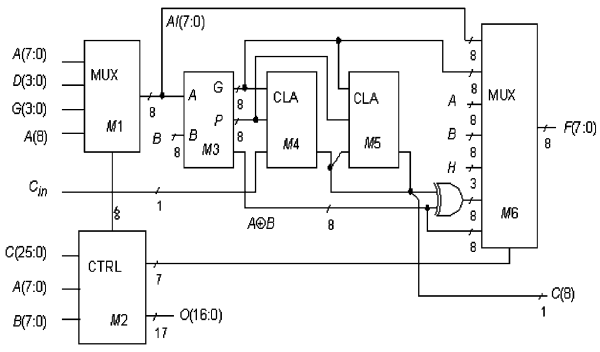
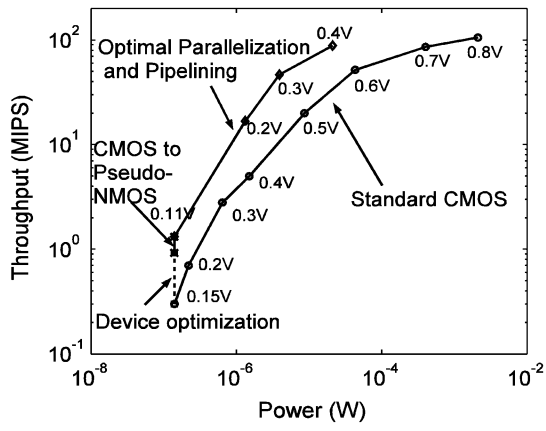


Fig. 13. Throughput versus power of a five-tap FIR filter with standard devices designed in CMOS style with optimal pipelining/parallelization (bottom line) and optimized devices in pseudo-NMOS style with optimal parallelization/pipelining (top line).



(a)



(b)

Fig. 14. (a) Schematic diagram of an 8-b ALU (c880) from ISCAS'85 benchmark. (b) Throughput versus power of the circuit in (a). Note how device optimization, choice of circuit style, and optimal parallelization/pipelining can dramatically improve throughput at iso-power.

is more than two times better (for iso-power) than that for the conventional design (Fig. 13).

The same technique has also been applied to an 8-b ALU (ISCAS'85 benchmark), shown in Fig. 14(a). As before,

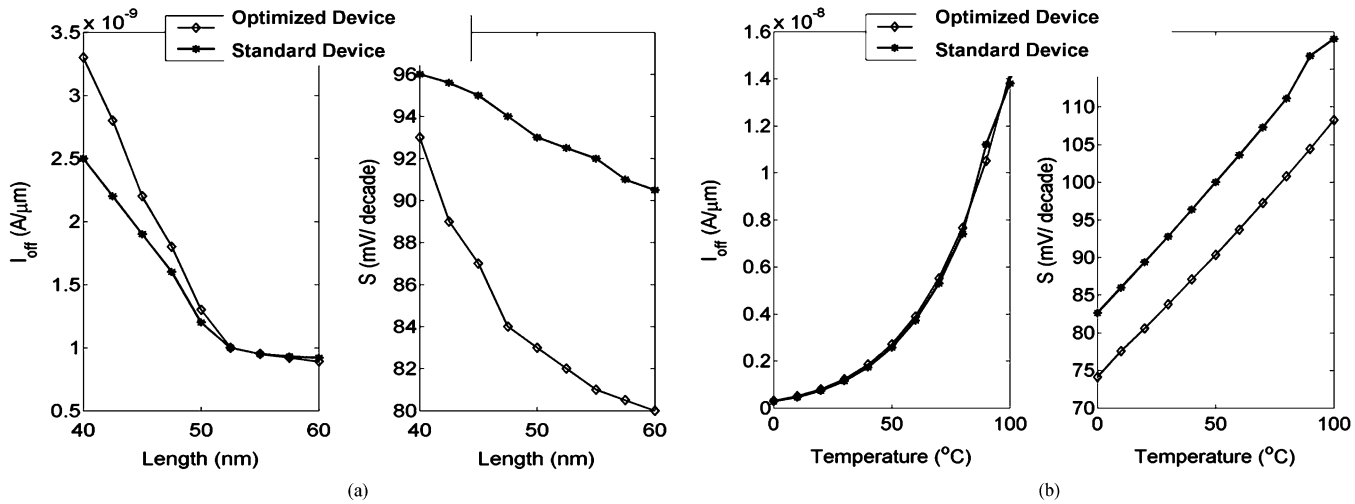


Fig. 15. Variation of $I_{\rm OFF}$ and S with (a) L and (b) temperature for the standard and the optimized devices.

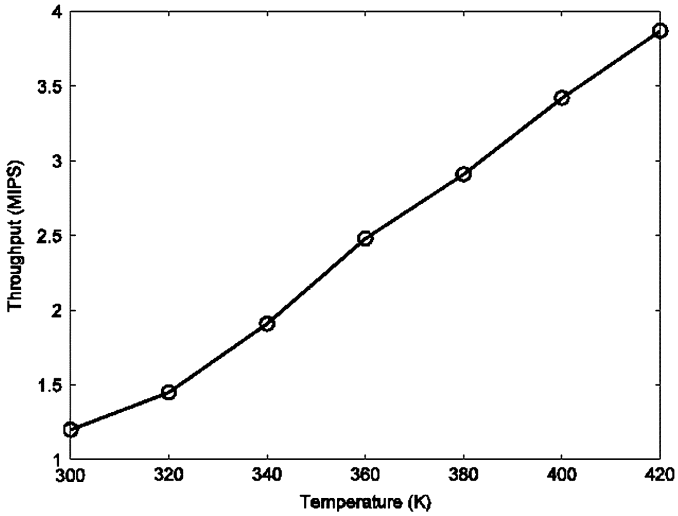


Fig. 16. Variation of throughput of the FIR filter (single pipeline stage and one parallel unit) with temperature.

we were able to achieve significantly higher throughput at iso-power by optimization at the device/circuit/architectural levels [Fig. 14(b)].

V. Effects of Variations and Noise

Like any super-threshold design, subthreshold logic is also affected by process, temperature and $V_{\rm dd}$ variations. Process variation is of primal concern in the device performance. Since we have optimized devices for better PDP, it is important to investigate its process tolerance. Fig. 15(a) illustrates the variation of $I_{\rm off}$ and S with channel length for the standard device and the optimized devices (at $V_{\rm dd} = 200$ mV) ¬. Note that the optimized device always has better S and has comparable $I_{\rm off}$ variation as the standard device. The impact of the operating temperature on $I_{\rm off}$ and S have been illustrated in Fig. 15(b). The optimized device and the standard device both show almost identical increase in the OFF current and a linear increase in S. Note that an increase in the operating temperature increases the operating current in subthreshold region and, hence, affects system performance favorably by making the circuits faster. Fig. 16 illustrates the variation of throughput of the previously described FIR filter with temperature. $V_{\rm dd}$ variations do play an important role in subthreshold domain since the current is exponentially dependent on $V_{\rm dd}$. Hence, a careful design of the power grid is critical. Then again, with smaller switching currents, Ldi/dt fluctuations are negligible in the subthreshold domain.

TABLE I
Worst Case Noise Margin (in Volts) of an Inverter in CMOS and Pseudo-NMOS Logic Styles

V_{dd} (V)	CMOS Logic Style		Pseudo NMOS Logic Style	
	Volts	Normalized w.r.t. V_{dd}	Volts	Normalized w.r.t. V_{dd}
0.8	0.35	0.437	0.18	0.225
0.6	0.26	0.450	0.124	0.233
0.4	0.18	0.450	0.13	0.325
0.2	0.09	0.450	0.07	0.350
The shaded cells represent subthreshold operation				

Another important aspect of the subthreshold domain of device operation is the reduced noise margin. Since the operating $V_{\rm dd}$ is low, both static (for CMOS) and dynamic (for DOMINO Logic) noise margins get affected. Further, use of the pseudo-NMOS logic family degrades the output low voltage and the corresponding noise margin. Table I shows how the worst case static noise margin changes as the $V_{\rm dd}$ is scaled down for an inverter. We have defined the noise margin as in [19]. For pseudo-NMOS, the noise margin corresponding to logic ZERO output is much lower. Hence, we have considered it as the worst case noise margin.

From Table I, it can be noted that the worst case noise margin degrades as the supply voltage is scaled down. This is often been identified as a critical bottleneck to design of ultralow-voltage circuits. However, a crucial thing to note is that, in the subthreshold region, the transconductance gain of the transistor increases. As a result, the voltage transfer characteristics become sharper. This can be noted from the noise margin expressed as a fraction of $V_{\rm dd}$. For CMOS logic, the noise margin (normalized with respect to $V_{\rm dd}$) increases from 0.42 to 0.45 whereas, for pseudo-NMOS, it increases from 0.21 to 0.35. Thus, although the absolute value of the noise margin decreases as the voltage is scaled down, the noise margin normalized with respect to the $V_{\rm dd}$ increases.

VI. Conclusion

This paper presents a design methodology in all levels of hierarchy (device, circuit, and architecture) for ultralow-power digital subthreshold operation. We have demonstrated that conventional design techniques are not optimal for subthreshold design. By proper co-design, it is possible to obtain hundreds of megahertz of performance in subthreshold systems while dissipating much lower consumption compared to its super-threshold (standard) counterpart.

References

[1] C. H. Choi, K. Nam, Z. Yu, and R. W. Dutton, "Impact of gate tunneling current in scaled MOS on circuit performance: A simulation study," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2823–2829, Dec. 2001.

[2] H. S. Momose, M. Ono, T. Yoshitomo, T. Ohguro, S. Nakamura, M. Saito, and H. Iwai, "1.5 nm direct-tunneling gate oxide Si MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, no. 8, pp. 1233–1242, Aug. 1996.

[3] J. Jomaah, G. Ghibaudo, and F. Balestra, "Band-to-band tunneling model of gate induced drain leakage current in silicon MOS transistors," *Electron. Lett.*, vol. 32, no. 8, pp. 767–769, 1996.

[4] C. Chen and M. Sarrafzadeh, "Simultaneous voltage scaling and gate sizing for low-power design," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 49, no. 6, pp. 400–408, Jun. 2002.

[5] C. H. Kim and K. Roy, "Dynamic VTH scaling scheme for active leakage power reduction," in *Proc. Design, Automation and Test in Europe Conf. Exhibit.*, 2002, pp. 163–167.

[6] A. Agarwal, H. Li, and K. Roy, "A single Vt low-leakage gated- ground cache for deep submicron," *IEEE J.Solid State Circuits*, vol. 38, no. , pp. 319–328, 2003.

[7] G. Palumbo, F. Pappalardo, and S. Sannella, "Evaluation on power reduction applying gated clock approaches," *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, pp. 85–88, 2002.

[8] Y. Cao, C. Hu, X. Huang, A. B. Kahng, S. Muddu, D. Stroobandt, and D. Sylvester, "Effects of global interconnect optimizations on performance estimation of deep submicron design," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, 2000, pp. 56–61.

[9] S. Katkoori and S. Alupoaei, "RT-level interconnect optimization in DSM regime," in *Proc. VLSI*, 2000, pp. 143–148.

[10] L. Entrena, C. Lopez, E. Olias, E. S. Millan, and J. A. Espejo, "Logic optimization of unidirectional circuits with structural methods," in *Proc. On-Line Testing Workshop*, 2001, pp. 43–47.

[11] E. S. Millan, L. Entrena, and J. A. Espejo, "On the optimization power of redundancy addition and removal for sequential logic optimization," in *Proc. Digital Systems Design*, 2001, pp. 292–299.

[12] J. Lim *et al.*, "Energy recovery logic circuit without nonadiabatic energy loss," *Electron. Lett.*, vol. 34, no. 4, pp. 344–346, Feb. 1998.

[13] H. Soeleman, K. Roy, and B. C. Paul, "Robust subthreshold logic for ultra-low power operation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 1, pp. 90–99, Jan. 2001.

[14] A. P. Chandrakasan, S. Sheng, and R. W. Broderson, "Low-power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, no. 4, pp. 473–484, Apr. 1992.

[15] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. New York: Cambridge Univ. Press, 1998.

[16] B. C. Paul, A. Raychowdhury, and K. Roy, "Device optimization for ultra-low power digital subthreshold operation," in *Proc. Int. Symp. Low Power Electronics and Design (ISLPED)*, Newport Beach, CA, Aug. 2004, pp. 96–101.

[17] Z. Lee, M. B. McIlrath, and D. A. Antoniadis, "Two-dimensional doping profile characterization of MOSFET's by inverse modeling using characteristics in the subthreshold Region," *IEEE Trans. Electron Devices*, no. 8, pp. 1640–1649, Aug. 1999.

[18] C. H.-I. Kim, H. Soeleman, and K. Roy, "Ultra-low-power DLMS adaptive filter for hearing aid applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 6, pp. 1058–1067, Dec. 2003.

[19] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, Dec. 2002.

Arijit Raychowdhury (M'02) received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Calcutta, India, in 2001. He is currently working toward the Ph.D. degree in electrical and computer engineering at Purdue University, West Lafayette, IN.

He has worked as an Analog Circuit Designer with Texas Instruments India. His research interests include device/circuit design for scaled silicon and nonsilicon devices.

Mr. Raychowdhury was the recipient of academic excellence awards in 1997, 2000, and 2001 and the Messner Fellowship from Purdue University in 2002. He was also the recipient of the "Best Student Paper" Award at the IEEE Nanotechnology Conference, 2003.

Bipul C. Paul (S'97–M'01–SM'05) received the B.Tech. and M.Tech. degrees in radiophysics and electronics from the University of Calcutta, Calcutta, India, and the Ph.D. degree from Indian Institute of Science, Bangalore, India.

After graduation, he joined Alliance Semiconductor, Bangalore, where he worked on synchronous DRAM design. In 2000, he joined Purdue University, West Lafayette, IN, as a Post-Doctoral Fellow, where he was involved with low-power electronic design of nanoscale circuits under process variation (using both bulk and SOI devices), testing, and noise analysis. He has also developed device and circuit optimization techniques for ultralow-power subthreshold operation. He is presently with Toshiba Corporation, CA, where he is working on postsilicon devices and technology.

Dr. Paul was the recipient of the Senior Research Fellowship Award from CSIR, India, in 1995 and the Best Thesis of the Year Award in 1999.

Swarup Bhunia (S'00–M'05) received the B.S. (Hons.) degree from Jadavpur University, Calcutta, India, the M.S. degree from the Indian Institute of Technology, Kharagpur, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 2005.

Currently, he is an Assistant Professor of electrical engineering and computer science with Case Western Reserve University, OH. He has worked in the EDA industry on RTL synthesis and verification for approximately three years. He has more than 40 publications in refereed journals and conferences. His research interests include design methodologies for high-performance, low-power testable VLSI systems, defect-based testing, noise analysis, and noise-aware design.

Prof. Bhunia was the recipient of the Best Paper Awards at the 2003 Latin American Test Workshop and the 2004 International Conference on Computer Design.

Kaushik Roy (S'83–M'90–SM'95–F'02) received the B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1990.

He was with the Semiconductor Process and Design Center of Texas Instruments, Dallas, TX, where he was involved with FPGA architecture development and low-power circuit design. He joined the Electrical and Computer Engineering Faculty, Purdue University, West Lafayette, IN, in 1993, where he is currently a Professor and holds the Roscoe H. George Professor of Electrical and Computer Engineering Chair. His research interests include VLSI design/computer-aided design for nanoscale silicon and nonsilicon technologies, low-power electronics for portable computing and wireless communications, VLSI testing and verification, and reconfigurable computing. He has published more than 300 papers in refereed journals and conferences, holds eight patents, and is a coauthor of two books on low-power CMOS VLSI design. He is the Chief Technical Advisor of Zenasis Inc. and a Research Visionary Board Member of Motorola Laboratories (2002). He was a Guest Editor for a special issue of the *IEE Proceedings Computers and Digital Techniques* (July 2002).

Dr. Roy was the recipient of the National Science Foundation Career Development Award in 1995, the IBM Faculty Partnership Award, the AT&T/Lucent Foundation Award, the 2005 SRC Technical Excellence Award, the SRC Inventors Award, and Best Paper Awards at the 1997 International Test Conference, the 2000 IEEE International Symposium on Quality of IC Design, and the 2005 IEEE Circuits and Systems Society Outstanding Young Author Award. He has been on the Editorial Board of *IEEE Design and Test,* the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS. He was the Guest Editor for a Special Issue on Low-Power VLSI of *IEEE Design and Test* (1994) and the IEEE TRANSACTIONS ON VLSI SYSTEMS (June 2000).