

Low-power circuits and technology for wireless digital systems

S. V. Kosonocky
A. J. Bhavnagarwala
K. Chin
G. D. Gristede
A.-M. Haen
W. Hwang
M. B. Ketchen
S. Kim
D. R. Knebel
K. W. Warren
V. Zyuban

As CMOS technology scales to deep-submicron dimensions, designers face new challenges in determining the proper balance between aggressive high-performance transistors and lower-performance transistors to optimize system power and performance for a given application. Determining this balance is crucial for battery-powered handheld devices in which transistor leakage and active power limit the available system performance. This paper explores these questions and describes circuit techniques for low-power communication systems which exploit the capabilities of advanced CMOS technology.

Introduction

As CMOS technology scales to deep-submicron dimensions, highly integrated system-on-a-chip (SoC) circuits can transform computing performance levels previously seen only on desktop computers to performance levels on small, handheld wireless devices. The capability for high-speed, low-power wireless computation opens up exciting new possibilities and applications. Wristwatch computers, such as the IBM WatchPad* or WorkPad*, or a cellular telephone with video streaming connected to a server through wireless LANs, are just a few examples. Simple scaling of conventional static CMOS circuits by technology alone will not produce the optimal low-power design. As we approach the limits of CMOS technology scaling, harnessing the computing performance made possible by these advanced silicon transistors in battery-operated devices becomes very challenging. For these, power becomes a main focus, and new circuit techniques and design methodologies are necessary to maximize the use of deep-submicron technology while maintaining an acceptable power consumption level. This paper begins by describing deep-submicron CMOS technology limitations, followed by low-power circuit techniques and multi-threshold design methodologies. Active well bias and power gating techniques are then described, followed by an in-depth study of latches and data retention, since this

is often a critical portion of overall power consumption. The final sections address architectural considerations, such as the use of parallelism, and power analysis of a third-generation (3G) mobile phone system.

Technology

CMOS technology continues to scale into deep-submicron levels, driven predominantly by performance requirements for desktop workstations and data servers. As MOS transistor channel lengths scale below 100 nanometers and device gate-oxide isolation scales below two nanometers, power consumption due to component leakage begins to surpass dynamic power. These parasitic leakage components are scaling faster than the dynamic switching power components [1, 2]. Technology scaling forces a lower bound on device threshold voltage due to increased subthreshold leakage. This lower bound causes traditional low-power techniques such as power-supply scaling to become less effective.

Figure 1 shows an inverter delay with a fan-out of 4 (FO4) for a bulk 0.13- μm CMOS technology as a function of the ratio of supply voltage (V_{dd}) and threshold voltage (V_t). As can be seen in the figure, voltage scaling can be used efficiently to trade off power for performance for V_{dd}/V_t ratios above 3, after which further voltage scaling becomes less efficient.

©Copyright 2003 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

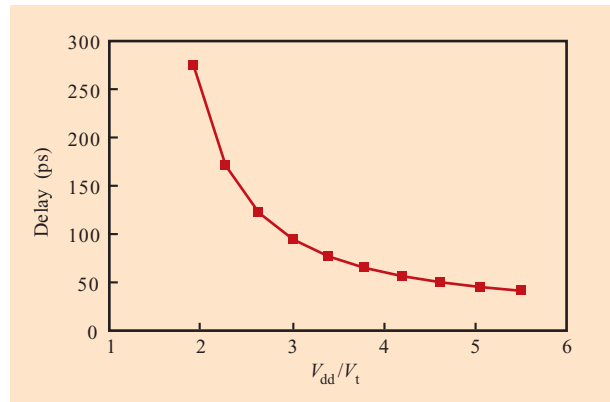


Figure 1

FO4 delay vs. V_{dd}/V_t ratio for a bulk 0.13- μm CMOS technology.

Low-power circuit styles

One approach to circumvent the technology limitations imposed on the circuit is optimizing the circuit style. Static CMOS is one of the most popular circuit styles for VLSI digital systems. This is due primarily to the robust design nature of this style, which can implement reliable circuits with the lowest sensitivity to process variations. Each digital system requires different power–delay and area tradeoffs. To meet these requirements, a number of different logic circuit families have been proposed [4, 5].

The power dissipation in CMOS circuits consists of several main components, as shown in Equation (1):

$$P_{\text{total}} = P_{\text{sw}} + P_{\text{sc}} + P_{\text{static}} + P_{\text{leak}}. \quad (1)$$

The dynamic switching power, P_{sw} , is the dominant active power dissipation component in CMOS circuits and results from the charging or discharging of the effective capacitive loads. The short-circuit power, P_{sc} , is due to the

existence of a conduction path between power supply and ground during the brief period when a gate switches, and it is usually small compared to P_{sw} . The static power component, P_{static} , is not usually a factor in pure CMOS circuit structures, but it is influenced by certain circuit structures such as sense amplifiers. The leakage power component, P_{leak} , is due to gate-oxide tunneling current, source-to-drain subthreshold conduction, and pn-junction leakage. These leakage components increase, becoming a larger percentage of total power dissipation, as transistor geometries shrink in future technology generations. The first two components of power dissipation, P_{sw} and P_{sc} , result from the actively changing states of the circuit. The last two components of power dissipation, P_{static} and P_{leak} , are always present and do not depend on the state changes of the circuit. The dynamic switching power, P_{sw} , is defined as

$$P_{\text{sw}} = \alpha C_L V_{dd}^2 f, \quad (2)$$

where α is the switching activity, C_L is the load capacitance, V_{dd} is the supply voltage, and f is the frequency of operation. Since the major active power component of a CMOS circuit is proportional to the square of the supply voltage, scaling of V_{dd} is a very effective way to reduce power dissipation. The leakage power, P_{leak} , is proportional to the device area and temperature. The subthreshold leakage component is strongly dependent on the device threshold voltage, V_t , and becomes an important factor as supply-voltage scaling is used to reduce power. Particularly for portable systems with a high ratio of standby to active operation, the leakage power may be the dominant factor in determining overall battery life. At present, advanced CMOS technologies offering two or more V_t choices can be effective in reducing leakage currents in the standby mode [5, 6].

Here we present a study of pass-transistor logic families and differential cascode voltage swing (DCVS) styles

Table 1 Relative three-input XOR ring oscillator simulation in a bulk 0.25- μm CMOS technology.

Gate	Delay	Power	Power $\times$ Delay	Area
CMOS	1.00	1.00	1.00	1.00
Differential cascode voltage switch w/pass-gate + buffer	0.52	0.79	0.41	0.41
Complementary pass-gate logic	0.55	1.51	0.85	0.38
Complementary pass-gate logic with 1/2 latch	0.62	1.21	0.68	0.54
Transmission pass-gate logic	0.54	1.24	0.68	0.7
Dual-layer pass-gate logic	0.52	0.76	0.40	0.4

compared to static CMOS circuits for low-power applications using single- V_t CMOS technology. In our experimental comparison, our test vehicle is a three-input XOR/XNOR logic gate. This logic function is very common in processor arithmetic circuits (e.g., adders, multipliers, or parity generators). The comparison methodology was studied in two test configurations to compare power, performance, and power-delay for these logic families. The first test configuration was a nine-stage ring oscillator in which transistor sizes were adjusted by a circuit-tuning computer-aided design tool to maximize the operating frequency of the oscillator. The second test configuration was a fixed-driver, variable-load configuration in which the sizes of the transistors were adjusted by the circuit-tuning tool to minimize the delay through the entire circuit from the input of the drivers to the output loads. The gates were tuned in this manner for a variety of output load strengths. All simulations were performed in 2.5-V, 0.25- μm CMOS technology. Shown in **Table 1** are results from a three-input ring oscillator experiment after frequency scaling and normalization. It was found that some logic styles (double-pass logic and differential cascode voltage switch with pass-gate and buffer) can have better power-delay than static CMOS cases [7]. The methodology of using a hybrid pass-transistor/differential cascode voltage switch (DCVS) and static CMOS circuit style has also been considered. By mixing static CMOS and pass-transistor logic families, the best overall power, speed, area, and energy-delay tradeoffs can be obtained.

Usage of multi-threshold devices

For portable systems with a high standby-to-active ratio, leakage current may be the dominant factor in determining overall battery life. We show how we can use mixed- or multi-threshold techniques to reduce subthreshold leakage current in the active mode of operation while maintaining the delay of the circuit near that of an all-low- V_t circuit. These logic circuits incorporate one set of low- V_t devices and another set of high- V_t devices. By appropriately selecting the low- V_t and high- V_t devices and their topological connections in the circuit, we can improve performance while minimizing leakage current. The key techniques are to use the low- V_t device to gain performance, high- V_t devices to cut off the leakage paths, and reverse-bias on the low- V_t devices in their standby state. The use of low- V_t devices can allow the designer to more effectively trade off performance for power savings by scaling voltage while maintaining performance. An algorithm for determining placement of low- V_t devices [8] produces an optimized mixture of low- V_t and high- V_t devices.

If there is no prior knowledge of the standby state input pattern or critical path, the general approach described

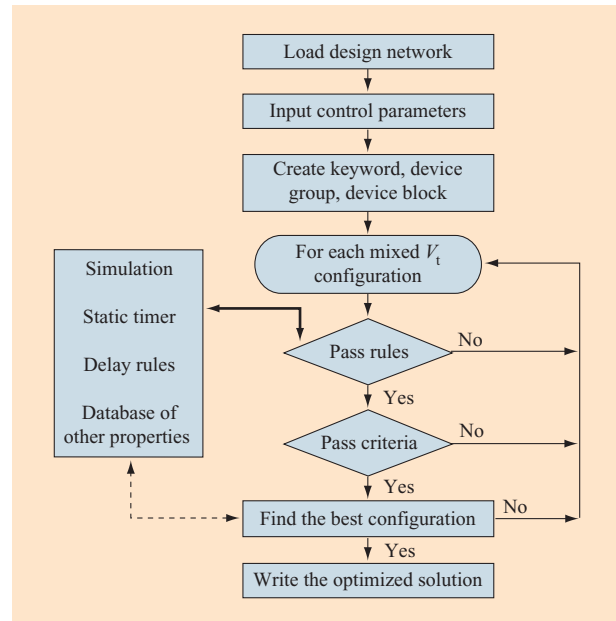


Figure 2

Flow chart for determining low- V_t devices. From [8], © 2001 ACM, Inc., reprinted by permission.

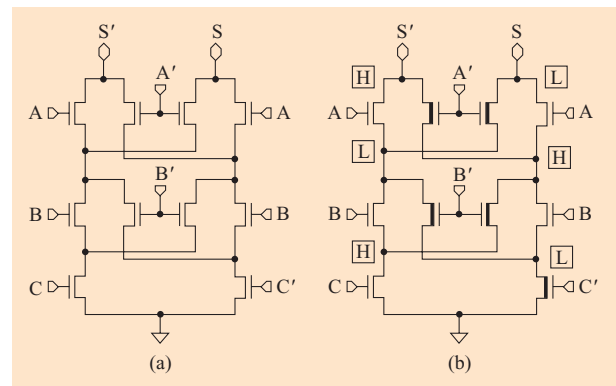


Figure 3

Evaluation-tree structures of one-bit adder sum in (a) regular- or low- V_t DCVS and (b) mixed- V_t DCVS.

in **Figure 2** can be used to select and compare different mixed- V_t structures. If, however, we have some knowledge of standby state input patterns, a much more effective method for selecting low- V_t device placement can be applied to the design of various logic circuits. Examples of mixed- and multi- V_t DCVS adder sum n-tree structures that are optimized by this method are shown in **Figures 3(a)** and **3(b)**. The correct logic behavior of the mixed- V_t

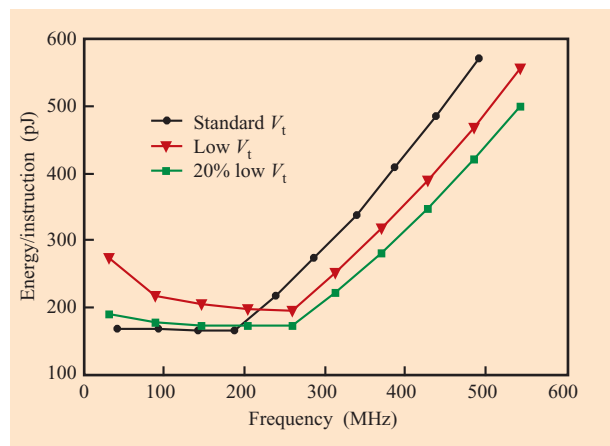


Figure 4

Average energy consumption per instruction for a general-purpose processor.

DCVS n-tree configuration has been confirmed by simulation results. The device model parameters used are based on multi- V_t 1.8-V 0.25- μm bulk CMOS technology. The delay, leakage current, and switching power simulation results for typical DCVS sum circuits with and without mixed- V_t devices in static, Domino CMOS, and NORA circuits are summarized in **Table 2**. The results show that the low- V_t DCVS circuits are the fastest of the cases considered.

Low- V_t DCVS standby leakage current, however, is the highest, about six times as much as that for high V_t (regular) and mixed V_t . The mixed- V_t DCVS circuits are almost as fast as their low- V_t counterparts, with a leakage current that is approximately as low as for the high- V_t DCVS cases. For applications which require low standby power, mixed- V_t DCVS is clearly the best choice. Also, as shown in Table 2, the Domino CMOS and NORA structures are much faster than the static CMOS circuits. The DCVS, Domino CMOS, and NORA circuits with multi- V_t show improved characteristics such as lower leakage and power, higher performance, and smaller layout area.

Figure 4 is a graph of the estimated average energy required to execute a general-purpose processor instruction at worst-case temperature and process over a range of operating frequencies. At each point, the supply voltage was adjusted from a range of 1.5 V to 0.6 V to just meet the frequency on the x -axis. It can be seen that the energy per instruction decreases at higher frequencies with lower V_t but increases at low frequencies because of the increased device leakage. It is also shown that by allowing 20% of the devices to use low- V_t thresholds, energy per operation can be optimized over a wide range of operating frequencies.

Active well bias

Another opportunity to obtain low-power CMOS circuits that is being increasingly studied is active well bias. Forward well bias of n-FETs (and corresponding reduced well bias of p-FETs) can be used to lower V_t and enhance performance when needed, while reverse well bias of n-FETs (and corresponding increased well bias of p-FETs) can be used to lower off-current in times of inactivity and reduced operating frequency. To give a better understanding of the performance implications of active well bias, we have designed and evaluated a set of 64 inverter ring oscillators in 0.18- μm , dual- V_t , triple-well CMOS technology.

The rings were configured so as to provide data for model-to-hardware correlation and also to provide direct useful information independently of any particular model. By varying the load driven by a given inverter design (i.e., different fan-outs, gate loads, diffusion capacitance loads, etc.) and measuring the power as well as the ring period, it is possible to decompose the delay into its resistive and capacitive components. Making such measurements as a function of well bias then provides insight into how these delay components change with well bias. **Figure 5** shows the percentage of change in stage delay as a function of well bias for six different circuit configurations from a “fast” chip with low V_t at $V_{dd} = 1.0$ V. Circuits 2, 3, 4, and 5 are all inverters with a fan-out of 3 having designs with various lengths of the polysilicon electrode, i.e., L_{poly} = nominal, nominal + 20 nm, nominal + 60 nm, and nominal + 90 nm, respectively. Although the delay

Table 2 Simulation results for the DCVS sum circuit.

	Delay (ps)			Leakage current (nA)			Switching power (pJ/switch)		
	Static	Domino	NORA	Static	Domino	NORA	Static	Domino	NORA
Regular V_t	107	75	67	35	6.7	6.7	0.64	0.353	0.37
Multi- V_t	85	68	62	36	7.2	7.2	0.64	0.366	0.38
Low V_t	81	66	61	214	40.0	40.0	0.63	0.372	0.39

increases with L_{poly} , the percentage of speedup with forward bias is greater for the longer-channel-length inverters, consistent with previously reported results [1]. Circuit 1 is similar to Circuit 2, except that Circuit 1 has a fan-out of 1. If all that changed with well bias was the drive capability of the inverter, the percentage of speedup with forward bias would be the same for Circuit 1 and Circuit 2. The fact that they are different is direct evidence that the capacitance also changes with well bias. To further illustrate this finding, Circuit 6 is an inverter that, in addition to the next inverter, also drives a diffusion capacitance load approximately equal to the input capacitance of two similar inverters. In this case, the percentage of increase in total capacitance with forward bias is actually greater than the percentage of increase in the drive capability, resulting in a net slowdown with forward bias. This data shows that the changes of both drive capability and capacitance with well bias are important, and that L_{poly} plays a major role. For any technology to take advantage of the performance opportunities as well as the directly related off-current control opportunities presented by active well bias, it is essential to understand and fully characterize the physical behavior, apply the technique to circuit situations on chip where the benefits can be maximized, and ensure that the device and circuit models accurately represent this behavior.

Power gating

Scaling the supply voltage and device threshold voltages (V_t) to a level which just maintains acceptable performance for a given application [9] has been shown to be an effective approach for low-power operation. This method achieves low active power, but comes at a cost of increased standby power resulting from increased device leakage from the use of low-threshold devices.

Multi-threshold CMOS (MTCMOS) [10] has been described as a method of reducing standby leakage current in the circuit by using a high-threshold MOS device to decouple the logic from the supply or ground during long idle periods, or sleep states. **Figure 6** shows an example of an MTCMOS circuit, where the logic block is constructed using low-threshold devices and either the power supply is gated by a high-threshold header switch, or the ground terminal is gated by a high-threshold footer switch.

During active operation of the MTCMOS circuit, the power-interrupt switch is turned on by the SLEEP' (or SLEEP) signal, and current dissipated by the logic is drawn through the interrupt switch, which causes a reduction in drive voltage seen by the logic, reducing logic performance. To compensate for the reduction in logic performance, larger power-supply voltages can be used at the expense of increased active power. For similar performance, larger device widths for the power interrupt

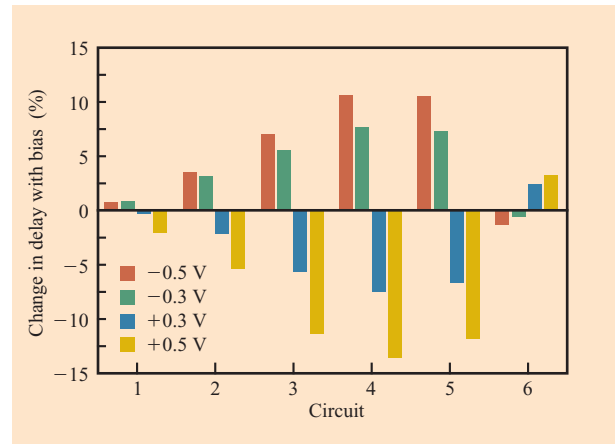


Figure 5

Percentage of change in delay with well bias for six different standard V_t inverter circuits on a "fast" chip at $V_{\text{dd}} = 1.0$ V: Circuit 1, standard inverter with fan-out of 1; Circuit 2, standard inverter with fan-out of 3; Circuit 3, standard inverter with fan-out of 3 and L_{poly} increased by 20 nm; Circuit 4, standard inverter with fan-out of 3 and L_{poly} increased by 60 nm; Circuit 5, standard inverter with fan-out of 3 and L_{poly} increased by 90 nm; and Circuit 6, standard inverter with fan-out of 1 and a large diffusion capacitance load.

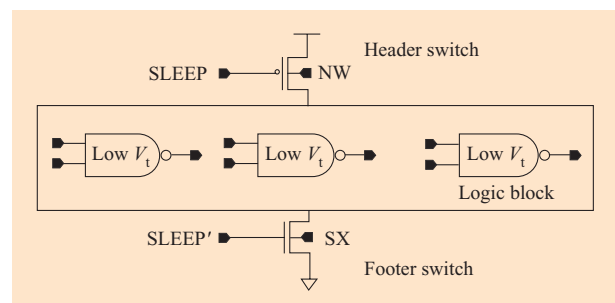


Figure 6

Example of MTCMOS circuit structure. From [11], © 2001 ACM, Inc., reprinted by permission.

switch can be used to minimize performance impact at the expense of increased area and power for entering and exiting sleep mode. Adjustment of device implants to allow moderately high-threshold values is another technique that can be used to increase performance at the expense of increased device leakage during idle mode. This section examines various available options for reducing frequency loss due to the insertion of a power-supply interrupt switch by means of gate drive and application of device body bias in an ASIC environment.

As a way to explore various options for implementing power-interrupt switches, a leakage model was constructed

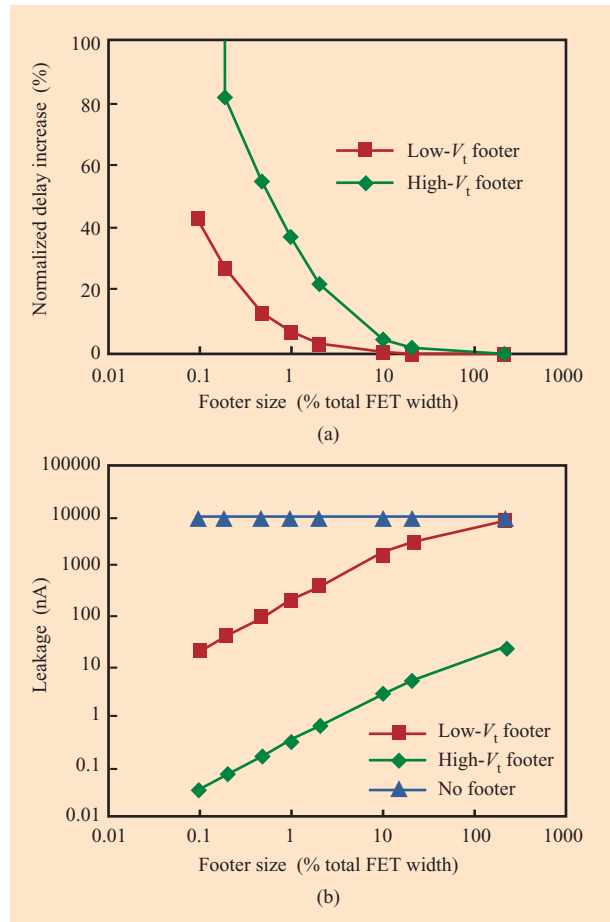


Figure 7

(a) Delay penalty and (b) total circuit leakage as a function of footer size at $V_{dd} = 0.9$ V and 85°C. From [11], © 2001 ACM, Inc., reprinted by permission.

to mimic the worst-case active power and leakage associated with the logic of a synthesized embedded processor [11]. Latches were omitted from the leakage model in order to simplify the experiment and to isolate effects on logic performance. These gates were arranged in eight delay chains spanning 2 ns to 4 ns in delay (under nominal conditions and supply voltage) to resemble various critical paths in the processor. The logic gates were placed and routed in an 0.18- μ m triple-well CMOS technology using low-threshold devices with isolated n-wells and p-wells. The threshold voltages defined in this experimental technology were approximately +0.25 and -0.25 V for the low- V_t n-FET and p-FET devices, respectively, and +0.7 V and -0.65 V for the high- V_t n-FET and p-FET devices, respectively.

Transistor-level circuit simulation was used to estimate the leakage of the test circuit along with the size

requirements of the header and footer switches. **Figure 7(a)** shows the normalized delay increase of the test circuit as a function of footer size as a percentage of total device width in the test circuit for a low- V_t footer and a high- V_t footer. The simulation was made at 0.9 V V_{dd} at 85°C. For both low- V_t and high- V_t -footer cases, the delay logic was made using only low- V_t devices.

Figure 7(b) shows leakage of the test circuit as a function of footer-device width in terms of percentage of total FET width in the test circuit for 0.9 V V_{dd} at 85°C. A curve for no footer shows a constant leakage which would result from the circuit without a footer, another curve shows the leakage with a low- V_t footer, and the third curve shows the leakage resulting from a high- V_t footer. Together, Figures 7(a) and 7(b), respectively, show the performance and leakage tradeoffs possible from choice of footer size and threshold voltage.

Various test circuits were constructed and fabricated with options for separately biasing the bodies of the power-interrupt switches with respect to the switched logic. All test measurements were performed at room temperature with a 0.9-V power supply. All gates in the processor model were constructed with low- V_t FETs. The footers were sized at 2% of the total n-FET and p-FET width of the processor model. Experiments incorporating a header device (V_{dd} interrupt) used a p-FET sized at 3.5% of the total n-FET and p-FET width. These sizings were chosen to minimize frequency impact and for equal footprint of n-FET footers and p-FET headers created by parallel instantiation of standard cell books.

Figure 8(a) shows the effects of n-well bias on a test macro incorporating an external ring connected to a high- V_t header. The test was run with a 0.9-V supply voltage. The plot also shows the performance improvement from forward bias of the header n-well. As can be seen from the plot, an additional performance increase of approximately 5% can be accomplished by forward-biasing the header n-well with respect to the switched logic n-well. The forward-bias effect is limited by increase in leakage current as the well diode begins to conduct in forward-bias mode.

Figure 8(b) shows the measured standby leakage of the test macro utilizing a low- V_t footer and a high- V_t footer with variable p-well bias. The test was also performed at room temperature and 0.9 V V_{dd} . During nominal p-well bias at 0 V, the high- V_t footer offers approximately 1000 $\times$ reduction in leakage, while under reverse bias of the footer switch, the leakage advantage from the increased V_t is diminished by the increased gate-induced drain leakage (GIDL) [12] seen in the high- V_t n-FET device. The figure also shows that a low- V_t footer device at -1 V p-well bias can achieve lower leakage than a high- V_t footer at the same well bias. Similar results have been obtained using p-FET headers, but with much less GIDL effect.

These results show that the use of a power-supply gating switch such as a header or footer device in series with low- V_t circuits can provide an effective means of decoupling inactive logic from the supply to reduce standby power. Circuit techniques such as gate overdrive and forward bias can be used to minimize the performance penalty of the power gate device.

Low-power latches

Reducing power dissipation in latches is essential for any low-power microprocessor. To identify latches suitable for low-power applications, we have conducted a comparative study of a number of different latch styles. Since the supply-voltage reduction is essential for reducing power, we focused primarily on static and semi-static latches because of their higher noise margin, and on those latches whose performance degrades the least as V_{dd} is reduced. It is important to consider the scan requirement at the early stages of the development of the clocking system, because the power overhead of the scannable design is significant; moreover, it is different for different latch styles.

A standard way of implementing a level-sensitive scan design (LSSD) master-slave latch uses transmission gates (called the PowerPC* latch or C²MOS latch in [13]). In these latches, the power overhead of the scan is quite small. However, these latches require two phases of clock (C for the master latch and B for the slave latch) in the normal operation mode. These two clock phases increase the total power of the clocking system by up to 30% for low-power latches using small transistor sizes. To avoid the power penalty of the second clock phase, the latch should operate with a single clock phase during the normal mode, and it should operate as a master-slave latch with two non-overlapping clock phases during the scan mode. To reduce the power overhead of the scan, it is desirable to separate the scan output of the latch from the data output, so that the wire connecting latches in the scan chain does not toggle in the normal operation mode.

Figures 9(a) and 9(b), respectively, show the developed extension to the modified sense amplifier [14–16] latch and the semi-static true single-phase RAM latch [17]. The result is achieved by mixing in the scan-in data at the second stage of the latch at the RS (reset/set) stage. The scan-in data signal I is written to the RS stage of the latch through transistors N1 and N2 or N3 and N4. The high level of clock A enables the scan-in write operation. The scan latches in Figures 9(a) and 9(b) are level-sensitive latches controlled by clock B. During scan mode, clock C is kept at the low level, and the RS stage of the SA latch and the scan latch work as a master-slave latch, controlled by clocks A and B. During normal operation mode, clocks A and B are kept at the low level, and the latch operates

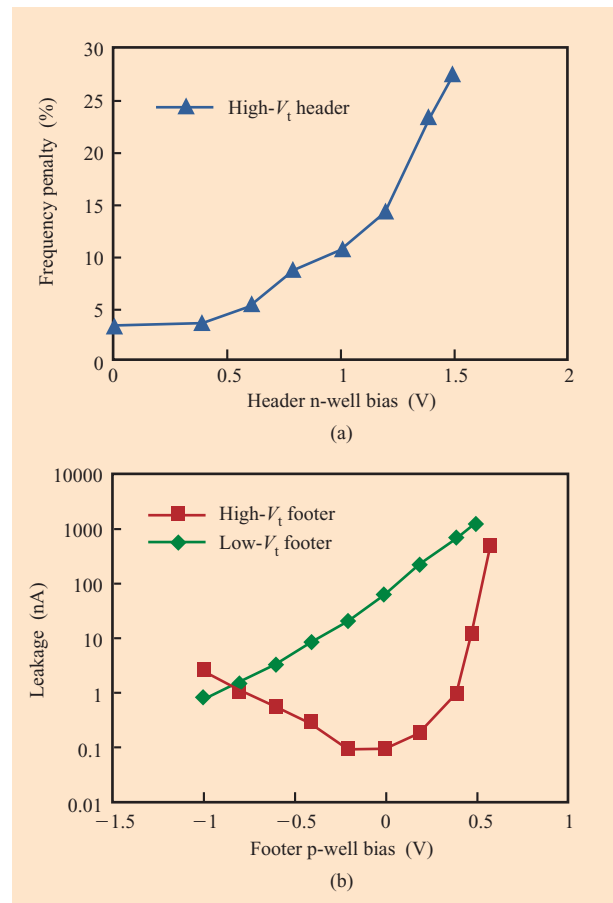


Figure 8

When $V_{dd} = 0.9$ V, (a) frequency penalty of high- V_t header with forward and reverse well bias; (b) leakage control with additional reverse p-well bias of either high- V_t or low- V_t footer. From [11], © 2001 ACM, Inc., reprinted by permission.

as the conventional sense-amp latch. The power overhead of the scan extension is $\Delta P = 0.5\alpha V_{dd}^2 C$, where C is the drain capacitance of transistors N1 and N3 plus the gate capacitance of two minimum-size transistors in the scan latch, and α is the “true” switching activity at the data input.

In **Figures 10(a) and 10(b)** we show the results of the comparative analysis of four latch styles: LSSD C²MOS (NORA) and transmission-gate (TG) latches, the developed scannable sense-amplifier (SA) latch of Figure 9(a), and the semi-static scannable true single-phase RAM latch of Figure 9(b). To fairly compare different latches in the power-performance space, we first built for every latch an energy-efficient family of configurations [16], which is a family of configurations obtained by tuning the latch to minimize the cost function $\gamma(E/E_0)^2 + (1 - \gamma)(D/D_0)^2$. The optimization parameter γ was varied in the range

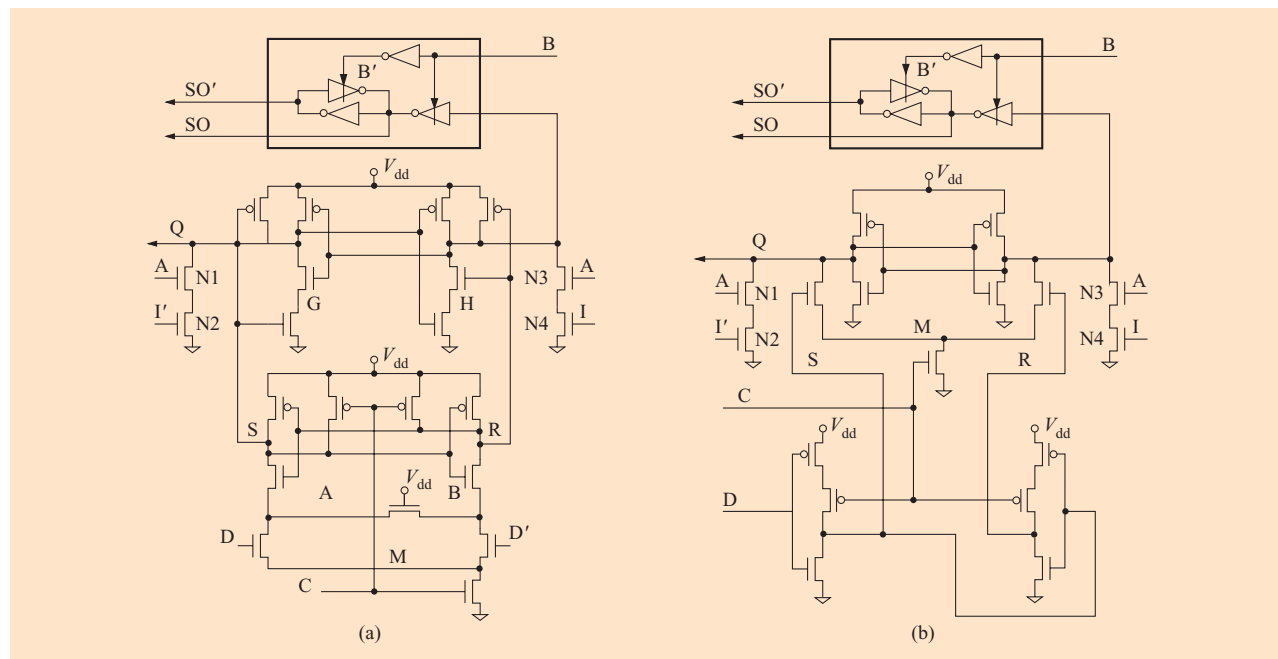


Figure 9

Transistor-level diagram of (a) modified sense amplifier (SA) latch; (b) semi-static true single-phase RAM latch.

from 0.1 to 0.9 V, and V_{dd} was set to 0.9 V. Here, D is the sum of the setup time and the delay through the latch; E is the average energy dissipated by the latch in one clock cycle simulated in a bulk CMOS 0.13- μm technology. E_0 and D_0 are estimates of the lower bounds on latch energy and delay.

The results are shown in Figure 10(a) for a switching activity factor of 0.3 transitions per cycle and a spurious activity of 0.15 glitches per cycle.¹ Extensive use of clock gating effectively increases the switching activity factor and the glitching activity factor. Figure 10(b) plots the average energy per cycle of the same latches for higher values of the switching and glitching activities. The low power consumption of the developed scannable sense-amplifier latch even in the presence of significant glitching activity, combined with its ability to operate with reduced swing signals, makes it a very good candidate for low-power designs for applications that can have the high switching activity that is typical of arithmetic circuits in digital signal processors.

Data retention

In low-power applications which employ the use of a power-gate device, all logic can be built with low-threshold transistors, with a high-threshold transistor serving as a

footer or a header. To maintain the state of the latches during sleep mode, a balloon latch is used which is built of high-threshold devices and connected to the real power and ground [18, 19]. According to [19], adding the balloon latch adds ten extra transistors to the flip-flop, increasing the transistor count from 16 to 26. It also leads to an increase in delay through the latch and active power, because of the extra parasitic capacitance of the two transistors that gate data to and from the balloon latch.

We developed a new, low-overhead data-retention scheme which combines the balloon latch with the scan mechanism [20]. **Figures 11(a)** and **11(b)** respectively show the high-level view of the sense-amplifier latch using the new data-retention scheme and the corresponding transistor-level diagram. The new flip-flop with retention uses the scan latch as a high-threshold storage for retaining data during sleep mode, and provides an extra path for restoring the data from the retention latch to the main flip-flop (transistors N3 and N4, N7 and N8). The retention latch and the clock B inverter are built either of high-threshold transistors or of regular transistors, which are capable of being back-biased. Real ground and V_{dd} are used as power terminals in the retention latch. The rest of the flip-flop is built of low-threshold transistors, and it may use either virtual V_{dd} or virtual ground to cut the leakage path during sleep mode.

¹ The term *spurious activity* refers to multiple transitions of the data in the signal caused by unbalanced propagation of signals through the logic preceding the latch.

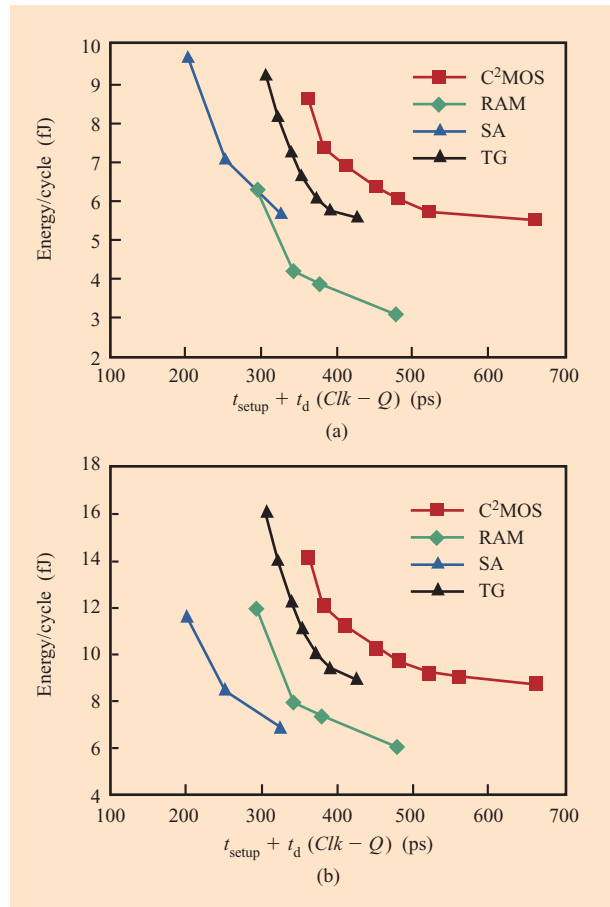


Figure 10

When $V_{dd} = 0.9$ V, average energy per cycle vs. performance: (a) switching activity factor = 0.3, glitching activity factor = 0.15; (b) switching activity factor = 0.5, glitching activity factor = 0.55.

During normal operation mode, clocks A and B are kept at the low level, and the latch operates as the conventional sense-amplifier latch. The RESTORE signal in Figure 11(a) [RSTR in Figure 11(b)] controls the multiplexing between the scan input I and the output of the retention latch (between the scan-in data and retention data). During scan mode, the RSTR is kept at the low level, and the latch works as a master-slave latch, controlled by clocks A and B, as described earlier. When entering sleep mode, a high level at clock B saves data in the retention latch. On returning from sleep mode, a high level is applied to the RSTR signal, and a high level at clock A restores data from the retention latch to the main flip-flop.

The power and delay overhead of the developed retention mechanism is reduced to a minor increase in capacitance of internal wires, due to some increase in

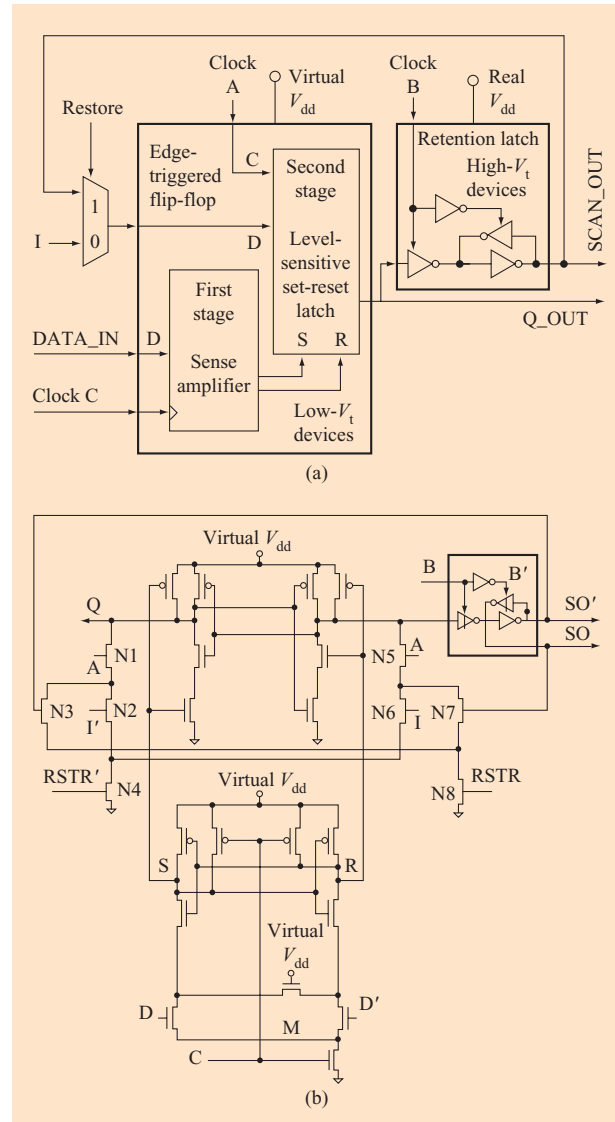


Figure 11

Modified sense-amplifier (SA) latch with data-retention scheme: (a) block diagram; (b) transistor-level schematic.

the area of the flip-flop (four extra n-FETs). No extra capacitance of transistor gates, sources, or drains is added to any nodes that are switching during the normal operation mode.

Parallelism

In this section, the opportunity to reduce power drain by exploiting concurrency-driven voltage scaling [21–23] is investigated at low supply voltages and deep-submicron feature sizes. The performance loss that accompanies scaling of supply voltage may be compensated for by

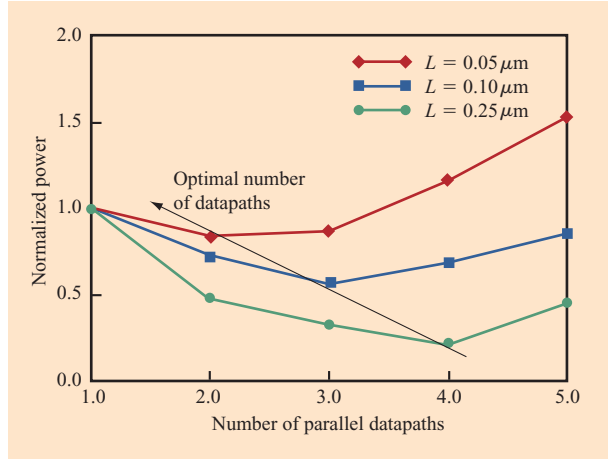


Figure 12

Normalized power dependence on number of parallel datapaths. Decreasing V_{dd}/V_t ratios across the roadmap yield smaller reductions in total power by exploiting datapath parallelism. From [22], © 2000 IEEE, reprinted by permission.

Table 3 Decreasing V_{dd}/V_t calculated from ITRS projections of supply voltage and saturation drain current per unit width.

Year	f_{clk} (GHz)	ITRS V_{dd}	V_{dd}/V_t required for I_{sat} of 600 $\mu A/\mu m$ (n-FET) and 280 $\mu A/\mu m$ (p-FET)
2001	1.4	1.5	3.0
2003	1.6	1.5	2.6
2006	2.0	1.2	2.55
2009	2.5	0.9	2.38
2012	3.0	0.6	2.43

adding parallel datapaths. This ensures that the total number of logic operations per second, F_o (system throughput), remains constant. Thus, for a prescribed throughput, total power drain from all of the parallel datapaths is lowered because of lower V_{dd} and a lower clock rate. However, increasing the number of parallel datapaths, N_p , increases the complexity and size of the overhead circuitry required for routing, multiplexing, and control of each of the parallel datapaths, and it increases static leakage, resulting in the dissipation of an additional component of overhead power. Also, latency increases by a factor N_p over the single-datapath case. A generic model for the dependence of the capacitance of overhead circuitry on the number of datapaths is described below and is used to compare the power drain of an optimized single datapath with that for parallel datapaths.

The clock requirements for each parallel datapath are reduced to

$$T_{cycle} = \frac{1}{f_{clk}} = \frac{N_p}{F_o}. \quad (3)$$

The power drain of N_p datapaths, each operating at the above clock rate, is given by

$$\begin{aligned} N_p P_{datapath} &= \alpha N_p C_{datapath} V_{dd}^2 \frac{F_o}{N_p} + P_{static} N_p + P_{static} N_p \\ &= \alpha C_{datapath} V_{dd}^2 F_o + P_{static} N_p, \end{aligned} \quad (4)$$

where $C_{datapath}$ is the total switching capacitance along the critical path of a datapath, P_{static} is the total static power dissipated by each datapath, and α is the circuit switch factor. The increase in the switching capacitance of the overhead circuitry with additional datapaths is calculated relative to the datapath capacitance using the following generic model:

$$\frac{C_{overhead}}{C_{datapath}} = m N_p^\omega + \Gamma. \quad (5)$$

The parameter m models the complexity of the control circuitry and/or any other component of the overhead that does not increase with each additional datapath. The ω exponent of N_p models the rate at which the routing and multiplexing requirements increase with each additional datapath and is specified by the size and complexity of the datapath. For an 8-bit adder/comparator datapath [24], data from layouts showed an approximately quadratic dependence on N_p in (5). Examples in [24] indicate a range on m from 0.1 to 0.7.

The dynamic power dissipated by the overhead circuitry and the datapaths is given by

$$P_{overhead}^{dynamic} = \alpha C_{overhead} V_{dd}^2 F_o = \alpha C_{datapath} \times m (N_p^2 - 1) V_{dd}^2 F_o. \quad (6)$$

Static power dissipated by the overhead circuitry is assumed to increase linearly with the overhead capacitance.

The sum total of datapath and overhead power is given by

$$\begin{aligned} P_{total} &= \alpha C_{datapath} V_{dd}^2 F_o \left[1 + \frac{C_{overhead}}{C_{datapath}} \right] + P_{static} \frac{C_{overhead}}{C_{datapath}} \\ &\quad + P_{static} N_p. \end{aligned} \quad (7)$$

This total power dissipation is normalized by dividing by N_g , the number of gates in each datapath, using $C_L = C_{datapath}/N_g$, where C_L is the load at the output of each datapath gate. The total power dissipation from all of the parallel datapaths, normalized by the number of gates in each datapath, is given by

$$P_{\text{total}}^{\text{normalized}} = \alpha C_L V_{\text{dd}}^2 F_0 \left[1 + \frac{C_{\text{overhead}}}{C_{\text{datapath}}} \right] + P_{\text{static}}^{\text{gate}} \left(\frac{C_{\text{overhead}}}{C_{\text{datapath}}} + N_p \right). \quad (8)$$

Figure 12 plots the above normalized power per gate for three generations in the projections from the International Technology Roadmap for Semiconductors (ITRS) shown in **Table 3**, showing smaller reductions in total power obtainable by employing parallel datapaths. In this particular example of an 8-bit datapath constructed with conventional circuit techniques, the benefits of parallelization diminish as technology scales in the future because of the decreased V_{dd}/V_t ratio for future technologies. Active well biasing and/or low- V_t devices with leakage control can be used to improve the results obtained by parallelism.

3G system analysis

System-level power control includes workload partitioning among system elements so that unneeded elements may be idled to reduce or eliminate active power. At the board level, idle component leakage is conveniently managed by removing power to unused integrated circuits. On-chip power-supply gating, as described in a previous section, may be used to provide similar function within an SoC [6, 11, 19]. Workload may also be partitioned into time intervals so that active system elements are periodically gated on and off. In this case, there is a tradeoff between the energy saved by reducing leakage current and the energy consumed by supply-gate switching. Another tradeoff consideration is the energy required to save and restore states of system elements that are to be periodically gated off.

An SoC with requirements for wide performance range and low power would satisfy the promises of 3G baseband and applications processing. Such an SoC should be capable of high-data-rate digital communication and high-performance application processing while maintaining conservative energy consumption for less-demanding uses. 3G cellular communications protocols support partitioning of tasks into variable-sized intervals so that applications, such as periodically checking for a base station instruction (standby mode), may be idle more of the time than active.

Figure 13(a) is a block diagram showing partitioning of a 3G handset processor for power-down applications. A low-voltage power supply and low-frequency clock are provided so that functions that are always on (real-time clock, interrupt control, and power management) are energy-efficient. Core logic functions are given a variable voltage supply and a variable speed clock so that voltage and frequency may be reduced when performance is not critical.

Table 4 lists the time intervals required to perform various power-up and power-down operations which

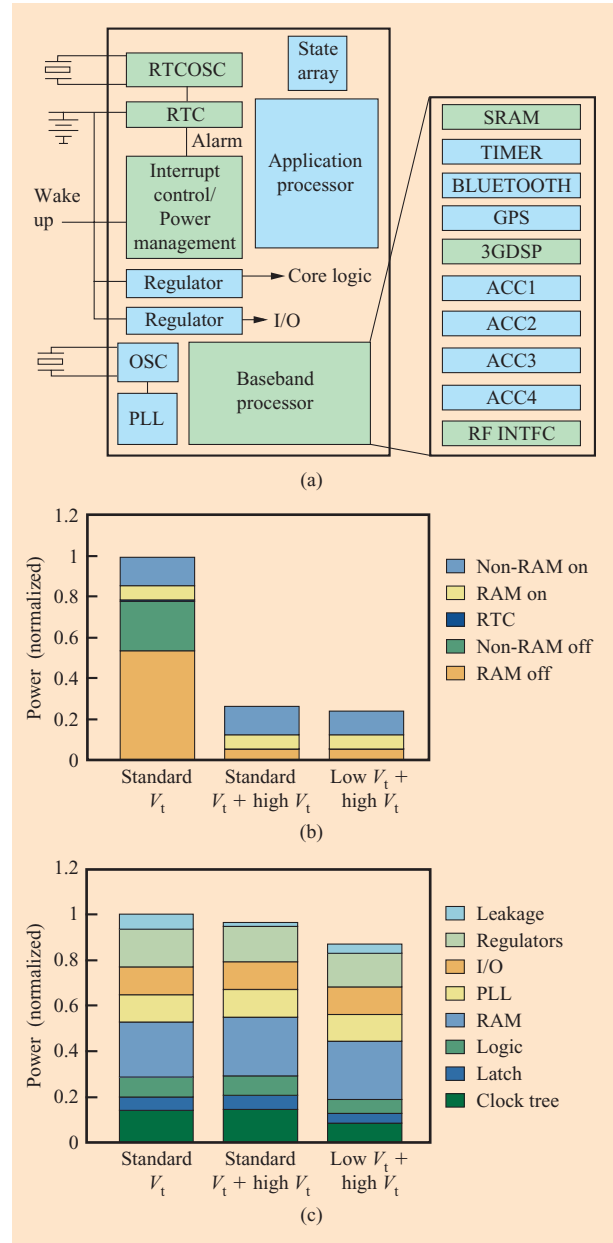


Figure 13

3G system analysis: (a) partitioning of an SoC for power-down applications (green blocks must be powered during phone standby operation, blue blocks are idle); (b) average power for standby waiting in call mode; (c) average power in active talk mode.

implement a temporal power management policy for a Wideband Code Division Multiple-Access (WCDMA) cellular communications protocol [25]. Such 3G protocols support partitioning of tasks into variable-sized intervals so that handset electronics may be idle more of the time than active. In this scenario the handset is in standby

Table 4 Power up/down times for WCDMA standby mode.

	Implementation 1	Implementation 2
Bring up PLL	100 μ s	100 μ s
Power up cores	1 μ s	1 μ s
Restore state	10 ms	0 s
Guard band	100 ms	100 μ s
Processing time	20 ms	20 ms
Save state	10 ms	0 s

mode, and the baseband processor elements of the SoC are activated for only brief periods to check for instructions transmitted from the nearest base station.

Since the base station must be polled every 1.28 seconds, implementation 1 may be gated off for 96.8% of the on/off cycle. It is obvious from the table entries that a large portion of the active mode is consumed by saving and restoring processor state. Power savings may be improved by implementing latches which retain state in the power-gated mode, as described in this paper in the section on data retention. Implementation 2 improves the off-state/on-state ratio to 98.4% by implementing these data-retention latches.

The core logic functions are separated into two major groups, one for baseband processing and another for application processing. The baseband processing functions include a low-power digital signal processor core capable of running a simple operating system independently of the applications processor. The application processing functions include a full-function microprocessor that takes control of the SoC only when the application demands it. Either major group may be completely powered down when not needed by the application.

Each processing group also includes a number of support processors (accelerators) for efficiently handling specific application needs. An example special-purpose accelerator function for 3G cellular communication is turbo-coding, which is required for digital data but not for voice. Supply gating is applied at this level of system partitioning so that current leakage suppression may be maximized without incurring a large die area penalty for the cutoff switches.

The efficiency of a leakage control implementation may be determined by comparing its power requirements to those of an implementation without leakage control. The addition of conventional MTCMOS cutoff transistors in series with the functional logic (power gating) requires that a higher supply voltage be applied to the circuit to achieve the same performance. This increases the active power of the circuit. If the energy reduction due to reduced leakage is less than the energy increase due

to higher active power plus the overhead of power gating, average power will increase.

Figure 13(b) shows the power breakdown of a potential 3G integrated SoC in standby mode, as described in Table 4, implementation 2. The bar charts show the power breakdown of each major circuit component compared to a baseline conventional ASIC design implemented with standard- V_t devices and without power gating devices for leakage control (bar 1). The middle bar shows the effect of adding high- V_t footer devices for leakage control, which reduces the average power in this mode by more than 74%. The third bar shows the marginal effect of replacing the standard- V_t devices in the active logic with low- V_t devices. Since very few logic circuits are active during a short period in this mode, power reduction from decreased supply voltage for active circuits does not greatly outweigh increased leakage resulting from the low- V_t devices.

Figure 13(c) shows the power breakdown for the same technology choices in active talk mode. In talk mode, idle circuit leakage is a smaller component of total power, which limits the amount of savings achievable through power gating. In this mode, active logic power and clock distribution power become more significant portions of average power and can benefit from the addition of low- V_t devices in the active logic, which allows a reduction in supply voltage as shown in the diagram. Many more modes can be defined by different application scenarios which exercise various portions of the chip and make extensive use of the application processor.

For the same die area, using higher-threshold transistors for power gating increases active power slightly but also dramatically reduces power consumption due to current leakage. The threshold of the transistor used in the active circuitry should also be adjusted to reduce total average power. Both active power and off-circuit power are best managed by using a lower-threshold transistor in the active circuit and a higher-threshold transistor in the cutoff switch, as shown in Figures 13(b) and 13(c). These examples demonstrate the value of a complete system view for analyzing power tradeoffs.

Conclusion

While advanced technology imposes new limitations on the circuit designer, integration and high-performance technology can enable new low-power applications if carefully designed. Leakage power is a major constraint for low-power battery-operated applications, but a combination of system architecture, circuit style, and technology options can allow long battery life while providing high performance on demand for new applications. Circuit designers of low-power battery-operated devices must make a key design choice for future

handheld systems: How to balance the need for low- V_t devices which allow voltage scaling for low active power and high frequency with the constraints of parallelism of function for high throughput and standby leakage power. The judicious use of high-performance devices to enhance performance at low supply voltages, high- V_t transistors to decrease leakage for noncritical paths and unused logic, and well bias enabled by triple-well structures to dynamically optimize V_t during run-time will help achieve performance while maintaining low power. Circuit choices such as tightly integrated designs, low-power circuit families, parallelism of function, application-specific hardware, low-power latch designs, and integrated power-supply-interrupt devices to deactivate idle logic and memory provide additional options for the circuit designer. Understanding all of these choices, their limitations, and how each of them scales with technology is critical to achieving an optimal design point for lower-power operation.

Acknowledgments

The authors wish to acknowledge the following individuals, whose insight, collaboration, technical exchange, and support made this work possible: from the IBM Microelectronics Division, Peter Cottrell, Terence Hook, Randy Mann, Jeff Brown, Edward Nowak, Tom Bednar, Jeffery Pauza, John Anderson, Roger Gregor (retired), Manjul Bhushan, and Richard Larsen; and from the IBM Research Division, Robert Dennard, Li-Kong Wang, Michael Immediato, Kevin Stawiaz, David Heidel, and David Meltzer (retired).

*Trademark or registered trademark of International Business Machines Corporation.

References

1. E. J. Nowak, "Maintaining the Benefits of CMOS Scaling When Scaling Bogs Down," *IBM J. Res. & Dev.* **46**, No. 2/3, 169–180 (March/May 2002).
2. J. Burr and A. Peterson, "Energy Considerations in Multichip-Module-Based Multiprocessors," *Proceedings of the International Conference on Computer Design*, September 1999, pp. 593–600.
3. K. Bernstein, K. Carrig, C. Durham, and P. Hansen, *High Speed CMOS Design Styles*, Kluwer Academic Publications, New York, 1998.
4. A. P. Chandrakasan and R. W. Brodersen, *Low Power Digital CMOS Design*, Kluwer Academic Publishers, Boston, 1995.
5. S. Kim, Y. Shin, S. Kosonocky, and W. Hwang, "Long-Term Power Optimization of Dual- V_t CMOS Circuits," *Proceedings of the 13th Annual IEEE International ASIC/SOC Conference*, September 2002, pp. 323–327.
6. J. T. Kao and A. P. Chandrakasan, "Dual-Threshold Voltage Techniques for Low-Power Digital Circuits," *IEEE J. Solid-State Circuits* **35**, No. 7, 1009–1018 (July 2000).
7. G. D. Gristede and W. Hwang, "A Comparison of Dual-Rail Pass Transistor Logic Families in 1.5V, 1.8 μ m CMOS Technology for Low Power Applications," *Proceedings of the 2000 Great Lakes Symposium on VLSI*, March 2000, pp. 101–106.
8. W. Chen, W. Hwang, P. Kudva, G. D. Gristede, S. Kosonocky, and R. V. Joshi, "Mixed Multi-Threshold Differential Cascode Voltage Switch (MT-DCVS) Circuit Styles and Strategies for Low Power VLSI Design," *Proceedings of the International Symposium on Low Power Electronics and Design*, August 2001, pp. 263–266.
9. I. Yang, C. Vieri, A. Chandrakasan, and D. Antoniadis, "Back Gated CMOS on SOI for Dynamic Threshold Control," *IEDM Tech. Digest*, pp. 877–880 (December 1995).
10. S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V Power Supply High-Speed Digital Circuit Technology with Multithreshold-Voltage CMOS," *IEEE J. Solid-State Circuits* **30**, No. 8, 847–854 (August 1995).
11. S. Kosonocky, M. Immediato, P. Cottrell, T. Hook, R. Mann, and J. Brown, "Enhanced Multi-Threshold (MTCMOS) Circuits Using Variable Well Bias," *Proceedings of the International Symposium on Low Power Electronics and Design*, August 2001, pp. 165–169.
12. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.
13. V. Stojanovic and V. Oklobdzija, "Comparative Analysis of Master-Slave Latches and Flip-Flops for High-Performance and Low-Power Systems," *IEEE J. Solid-State Circuits* **34**, No. 4, 536–548 (April 1999).
14. N. Nedovic and V. Oklobdzija, "Dynamic Flip-Flop with Improved Power," *Proceedings of the International Conference on Computer Design*, June 2000, pp. 323–326.
15. B. Nikolic, V. Oklobdzija, V. Stojanovic, W. Jia, J. Chiu, and M. Leung, "Improved Sense-Amplifier-Based Flip-Flop: Design and Measurements," *IEEE J. Solid-State Circuits* **35**, No. 6, 876–883 (June 2000).
16. V. Zyuban and D. Meltzer, "Clocking Strategies and Scannable Latches for Low Power Applications," *Proceedings of the International Symposium on Low Power Electronics and Design*, August 2001, pp. 346–351.
17. C. Svensson and J. Yuan, "Latches and Flip-Flops for Low Power Systems," in A. P. Chandrakasan and R. W. Brodersen, *Low-Power CMOS Design*, IEEE Press, New York, 1998, pp. 233–238.
18. S. Mutch, S. Shigematsu, Y. Matsuya, H. Fukuda, and J. Yamada, "A 1V Multi-Threshold Voltage CMOS DSP with an Efficient Power Management Technique for Mobile Phone Applications," *Proceedings of the International Solid-State Circuits Conference*, 1996, pp. 168–169.
19. S. Shigematsu, S. Mutoh, Y. Matsuya, Y. Tanabe, and J. Yamada, "A 1-V High-Speed MTCMOS Circuit Scheme for Power-Down Application Circuits," *IEEE J. Solid-State Circuits* **32**, No. 6, 861–869 (June 1997).
20. V. Zyuban and S. Kosonocky, "Low Power Integrated Scan-Retention Mechanism," *Proceedings of the International Symposium on Low Power Electronics and Design*, August 2002, pp. 98–102.
21. A. Chandrakasan, S. Sheng, and R. Brodersen, "Low Power CMOS Digital Design," *IEEE J. Solid-State Circuits* **27**, No. 4, 473–484 (April 1992).
22. A. Bhavnagarwala, B. Austin, K. Bowman, and J. Meindl, "A Minimum Total Power Methodology for Projecting Limits on CMOS GSI," *IEEE Trans. VLSI Syst.* **8**, No. 3, 235–251 (June 2000).

23. A. Bhavnagarwala, B. Austin, A. Kapoor, and J. Meindl, "CMOS System-on-a-Chip Voltage Scaling Beyond 50nm," *Proceedings of the 10th Great Lakes Symposium on VLSI*, March 2000, pp. 7–12.
24. A. P. Chandrakasan and R. W. Brodersen, *Low-Power CMOS Design*, IEEE Press, New York, 1998.
25. J. Korhonen, *Introduction to 3G Mobile Communications*, Artech House Inc., Norwood, MA, 2001.

Received February 12, 2002; accepted for publication October 24, 2002

Stephen V. Kosonocky *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (stevekos@us.ibm.com).* Dr. Kosonocky received B.S., M.S., and Ph.D. degrees from Rutgers University in 1986, 1991, and 1994. From 1986 to 1992 he was a Research Scientist at Siemens Corporate Research in Princeton, New Jersey, where he worked on digital and analog circuit design. From 1992 to 1993 he was at the Samsung Princeton Design Center, where he worked on mixed-signal video circuits. In 1994 he joined the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, where he has worked on embedded DRAM systems-on-a-chip, circuit, logic, and microarchitecture design of Daisy and Boa VLIW processors. Since 2000 he has been leading a team on low-power circuits, where he is interested in developing ultralow-power signal processing cores for communications systems and low-power embedded PowerPC processors.

Azeez J. Bhavnagarwala *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (azeezb@us.ibm.com).* Dr. Bhavnagarwala received his B.S. degree (cum laude) in 1992 from Rensselaer Polytechnic Institute (RPI) and from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland; his M.S. degree from RPI in 1994; and his Ph.D. degree from the Georgia Institute of Technology in 2001, all in electrical engineering. He was with Integrated Device Technology in 1996 and with LSI Logic in 1999, working on interconnect and device modeling. Dr. Bhavnagarwala's research interests are voltage scaling constraints for CMOS technologies. He has published more than 20 technical papers in conferences and journals. He received a Best Paper in Session award from the Semiconductor Research Corporation in 2000 and serves on the Technical Program Committee of the IEEE International ASIC/SoC Conference.

Kenneth Chin *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (kchin@us.ibm.com).* Mr. Chin received his B.S.E.E. and M.S.E.E. degrees from the New Jersey Institute of Technology in 1974 and 1979, respectively. From 1974 to 1977 he was with the Monsanto Company, Havre-de-Grace, Maryland, working on process instrumentation for high-speed manufacturing. In 1979 he joined the Microelectronics Division of General Instrument, Hicksville, New York, where he was involved in the design of microcomputer and nonvolatile chips. He joined the IBM General Technology Division in 1981, working in commercial SRAM qualification. From 1985 to 2000 he was with the High-Performance Circuit Design group at the IBM Thomas J. Watson Research Center, working on high-speed circuit design. From 1994 to 2000 he was part of the team that worked on the first three generations of IBM CMOS mainframe computers. In 2000 Mr. Chin joined the Low Power Circuits and Technology group. He has authored or co-authored more than 25 technical papers and is currently pursuing the Ph.D. degree in electrical engineering at Polytechnic University, New York.

George D. Gristede *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (gristede@us.ibm.com).* Dr. Gristede received a B.A. degree in professional option engineering from Columbia College in 1984. In 1985, 1988, 1990, and 1992, respectively, he received the B.S., M.S., M.Phil., and Ph.D. degrees from Columbia University, all in electrical engineering. His Ph.D.

thesis dealt with the derivation of a new mathematical formulation for analyzing the convergence properties of relaxation-based simulation methods. In 1992 Dr. Gristede joined the IBM Thomas J. Watson Research Center as a Research Staff Member; he has done extensive work in the areas of low-power, high-performance circuit design and the development of a new generation of custom CAD tools and the methodology to support such designs. His research interests include low-power, high-performance circuit design, automated circuit checking, circuit optimization, and automated physical design and optimization. Dr. Gristede is a member of Tau Beta Pi and Eta Kappa Nu.

Anne-Marie Haen *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ahaen@us.ibm.com).* Ms. Haen received the engineering diploma from the National Superior Engineering School of Electrical Engineering and Radio-Electricity (ENSERG), Grenoble, France in 1989. She joined IBM France in 1989 to develop custom embedded CMOS SRAMs, and, subsequently, custom digital circuits and FPGAs for asynchronous transmission mode (ATM) broadband network and multiprotocol network switches. In 1997, Ms. Haen joined the VLSI Design Department at the IBM Thomas J. Watson Research Center to work on high-performance CMOS microprocessors. Since 2000 she has been in the Low Power Circuits and Technology group, concentrating on the design system and tools for low-power design.

Wei Hwang *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (whwang@us.ibm.com).* Dr. Hwang has been a Research Staff Member at the IBM Thomas J. Watson Research Center since 1984. Prior to this, he was an Associate Professor in the Electrical Engineering Department at Columbia University from 1979 to 1984. He was an Assistant Professor of Electrical Engineering at Concordia University, Montreal, Canada, from 1975 to 1978. In 1974 he received the Ph.D. degree from the University of Manitoba, Canada. Dr. Hwang's interests are in the area of VLSI circuits and technology, DRAM and SRAM chip design, high-frequency server microprocessors, and rf microelectronics. His current research interests are in ultralow-power circuits and technology, elite digital signal processor (DSP) core design, and system-on-a-chip (SoC) design. He has received several IBM awards, including sixteen Invention Achievement Awards and four Research Division Technical Awards, and has been elected an IBM Master Inventor. Dr. Hwang is the co-author of the book *Electrical Transport in Solids*, which has been translated into Russian and Chinese. He has also published more than 100 technical papers in journals and conferences and holds 53 U.S. patents. Dr. Hwang is a Fellow of the Institute of Electrical and Electronics Engineers.

Mark B. Ketchen *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (mketchen@us.ibm.com).* Dr. Ketchen did his undergraduate work in physics at MIT and holds a Ph.D. degree in physics from the University of California at Berkeley. He served for four years as an officer in the U.S. Navy, teaching thermodynamics, nuclear physics, and nuclear reactor operations in the Naval Nuclear Power Program. Over the last 24 years he has held a variety of research and research management positions in solid-state science and technology at IBM. In 1998 he completed a three-year

assignment as Director of Physical Sciences at the IBM Thomas J. Watson Research Center, a job which included responsibility for IBM strategy in the physical sciences at the company's research laboratories in Yorktown Heights, New York, Almaden, California, and Zurich, Switzerland. Dr. Ketchen is currently Manager of the Theory and Computation group in the Physical Sciences Department. He is an expert in the area of microelectronic devices and measurement techniques, and is currently focusing on the use of computation to help define the trajectory of CMOS technology to the limits of scaling. Dr. Ketchen is a Fellow of the IEEE, a Fellow of the American Physical Society, a Member of the IBM Academy of Technology, and the recipient of the 1995–1996 American Institute of Physics Prize for Industrial Applications of Physics and the 1996 IEEE Morris E. Leeds Award.

Suhwan Kim *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (suhwa@us.ibm.com).* Dr. Kim received the Ph.D. degree in electrical engineering and computer science from the University of Michigan at Ann Arbor in 2001, joining the IBM Thomas J. Watson Research Center that same year. He is currently a Research Staff Member. Dr. Kim's research interests encompass high-performance and low-power circuits/microarchitecture and low-power design methodologies for high-performance VLSI signal processing. He received the 1994 Best Student Paper Award in the IEEE Korea Section and the First Prize in the VLSI Design Contest of the 2001 ACM/IEEE Design Automation Conference. Dr. Kim has participated several times in the Technical Program Committee of the IEEE International ASIC/SoC Conference.

Daniel R. Knebel *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (knebel@us.ibm.com).* Mr. Knebel is a Senior Engineer in the Digital Communication Engines Department at the IBM Thomas J. Watson Research Center. He joined the IBM Data Systems Division in 1982 to research and develop high-performance LSI circuits for System/370. He has since worked in a variety of related areas, including microprocessor design, VLSI test and diagnostics, electronic design automation, and, recently, low-power circuit research. Mr. Knebel received a B.S. degree in electrical engineering from Purdue University and has various publications, inventions, and patents pending related to his research and development activities at IBM.

Kevin W. Warren *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (kwwarren@us.ibm.com).* Dr. Warren received his Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1990. From 1990 to 1992, he held a research and teaching faculty position at the University of Illinois in the Department of Electrical and Computer Engineering. In addition, from 1986 to 1992 he helped found and served as a consultant for a flat-panel display manufacturing company. He joined the IBM Thomas J. Watson Research Center in 1992 and has held a variety of positions, including Manager of the Flat Panel Display Systems Group and Technical Assistant to the Vice President. Dr. Warren is now Senior Manager of the Digital Communications Engines Department, which comprises team members from the Haifa, Yorktown, and Zurich Research laboratories and from the Microelectronics Division. The team is responsible for research and development of all technology

associated with the digital aspects of network access. This includes all technologies from low-power technologies, circuits, and embedded architectures, to next-generation DSP technology, to signal processing application software.

Victor Zyuban *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (zyuban@us.ibm.com).* Dr. Zyuban received his B.S. and M.S. degrees from the Moscow Institute of Physics and Technology in 1993 and 1995, respectively, and his Ph.D. degree in computer science and engineering from the University of Notre Dame in 2000. From 1995 to 1996 he worked in the Moscow Center for SPARC Technologies. He is currently a Research Staff Member at the IBM Thomas J. Watson Research Center. Dr. Zyuban is working on a low-power DSP research project in which he has been involved in ISA definition, microarchitecture, and physical design. He is currently leading the development of a semicustom eLite core test chip. His research interests include low-power circuitry, microarchitecture, and methodologies for low-power design.