

Architecture and Circuit Techniques for Low-Throughput, Energy-Constrained Systems Across Technology Generations

Mark Hempstead, Gu-Yeon Wei, David Brooks
Division of Engineering and Applied Sciences, Harvard University
Cambridge MA, 02138

{mhempste, guyeon, dbrooks}@eecs.harvard.edu

Abstract

Rising interest in the applications of wireless sensor networks has spurred research in the development of computing systems for low-throughput, energy-constrained applications. Unlike traditional performance oriented applications, sensor network nodes are primarily constrained by operation lifetime, which is limited by power consumption. Advanced CMOS process technologies provide ever increasing transistor density and improved performance characteristics. However, shrinking feature size and decreasing threshold voltages also lead to significant increases in leakage current, which is especially troublesome for applications with significant idle times.

This work investigates tradeoffs between leakage and active power for low-throughput applications. We study these issues across a range of process technologies on a computing architecture that provides explicit support for fine-grain leakage-control techniques such as Vdd-gating and adaptive body bias. We present a methodology for selecting design parameters, including choice of process technology, that makes the optimal tradeoff between active power and leakage power for a given workload. Our results show that leakage power will dominate the selection of process technology, and architectures that support advanced leakage control techniques at the circuit level will be essential. We argue that without advanced low-power architectures future nano-scale process technologies will not be suited for sensor network applications.

Categories and Subject Descriptors

C.3 [Special-Purpose and Application-Based Systems]: Real-time and embedded systems; B.7.1 [Integrated Circuits]: Types and Design Styles—*Advanced technologies*; C.0 [General]: Modeling of computer architecture

General Terms

Design

Keywords

Wireless Sensor Networks, Technology Scaling, Low Power, System Architecture, Leakage Power Reduction

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CASES'06, October 23–25, 2006, Seoul, Korea.

Copyright 2006 ACM 1-59593-543-6/06/0010 ...\$5.00.

1. INTRODUCTION

Aggressive technology scaling has fueled explosive growth in the semiconductor industry, enabling ever faster and more powerful computational devices. At the other end of the spectrum, there has been growing interest in very low throughput devices, such as SoCs for wireless sensor networks, where low-power operation is the primary design target. Sensor networks have been proposed and deployed for a wide variety of applications such as habitat monitoring [19, 28], structural monitoring, and emergency medical response [12, 18]. Typical sensor network workloads require low-throughput devices that have low real-time computation requirements. A characteristic monitoring application remains idle most of the time and wakes up periodically to take sensor samples, send the sampled data using a radio, and then go back to sleep. The sampling rate depends on the phenomena being observed but is often on the order of 0.02Hz to 100Hz [19, 32]. While the application space seems limitless, the operating lifetime of the battery-operated wireless sensor nodes is a major limitation. Ideally, system developers would like to embed sensor network devices into the environment, such as into the walls of a new building. The deployed nodes would scavenge energy from the ambient environment [26] to report on the health of the structure over the lifetime of the building.

As widespread deployments of wireless sensor networks become more prevalent, we expect to see the development of computing devices that specifically target these low-throughput applications. There have been several early system implementations targeting this class of applications [7, 13, 22, 23]. In anticipation of this growing area, this paper provides insights into the benefits of architectures that support aggressive leakage control techniques in the presence of current trends in process technology scaling. Because sensor nodes are often idle for extended periods of time, designers must carefully tradeoff active and leakage current, as the latter can be the dominating contributor to overall energy consumption. This is especially true as recent generations of CMOS process technology have seen drastic increases in static leakage current due to reduction of threshold voltages and shrinking of gate oxides.

This paper presents simulation results using Berkeley Predictive Technology Models (BPTM) [4], to highlight trends in power consumption across technology generations. Moreover, we investigate the efficacy of several circuit techniques currently employed to decrease leakage current. We analyze the system-level benefits of these techniques in the context of a sensor node architecture that provides explicit support to utilize these leakage control techniques. We conclude that while these techniques can provide significant energy savings, more advanced CMOS technologies are not necessarily the best choice for low-power and low-throughput applications such as wireless sensor networks.

Section 2 presents a background of power dissipation in CMOS technology including several leakage-reduction techniques. Section 3 outlines our sensor node architecture and discusses explicit support for active and leakage power savings. Section 4 introduces our simulation framework and presents results in the context of a simple circuit. We connect the results from this framework with our system level model in Section 5, and we present data and an analytical model that provides system designers insights into the benefits of advanced process technologies and leakage control techniques when building devices for low-throughput applications.

2. BACKGROUND

2.1 Power Models

The total power consumption of a CMOS circuit consists of active plus leakage power (ignoring short-circuit current). Usually, the primary source of power consumption is active power, which can be modeled by:

$$P_{active} = \alpha CV^2 f \quad (1)$$

where α is the activity factor, C is the total switched capacitance, V is the supply voltage (and voltage swing), and f is the operating frequency. A well-designed system can take advantage of the low-throughput nature of sensor network applications to substantially reduce active power by reducing the activity, frequency, and voltage.

Another component of power is leakage current I_{leak} , which has been studied and modeled by several researchers [16, 21, 25]. To provide system designers and architects in the early stage of design insights into factors that affect leakage, we present a simplified BSIM model as described in [9]. Two significant components of leakage current (I_{leak}) for advanced CMOS technologies is subthreshold conduction (I_{sub}) and gate-oxide leakage (I_{ox}).

Subthreshold leakage current depends on the threshold and supply voltages.

$$I_{sub} = K_1 W e^{-V_{th}/nV_{\theta}} (1 - e^{-V/V_{\theta}}) \quad (2)$$

where K_1 and n are experimentally derived, W is the gate width, and V_{θ} is the thermal voltage. Notice that linear scaling of V_{th} causes subthreshold leakage to increase exponentially.

Gate-oxide leakage has become significant for advanced technology generations (130nm and beyond). The following is a simplified model for gate leakage current that was presented in [9].

$$I_{ox} = K_2 W \left(\frac{V}{T_{ox}}\right)^2 e^{-\alpha T_{ox}/V} \quad (3)$$

where K_2 and α are experimentally derived and T_{ox} is the thickness of the gate oxide. This relationship shows that decreasing T_{ox} exacerbates gate leakage. However, T_{ox} generally follows scaling trends to avoid short channel effects.

In addition to these sources of leakage power, the BSIM models [6] include other sources of leakage such as junction leakage, drain-induced barrier lowering current (DIBL) and gate-induced drain leakage (GIDL). It is important to note that temperature also has a significant impact on leakage although sensor nodes generally will operate at relatively low temperatures (less than $40^{\circ}C$).

2.2 Technology Scaling

The scaling of CMOS technology across process generations has followed several well documented trends [2, 11]. These (constant field) scaling trends include:

- 30% gate delay reduction, 43% operating frequency increase
- Active energy/cycle scales down by 70% per generation
- V_{th} decreases by 15% causing a $5\times$ increase in subthreshold current per generation

These scaling trends are projected to continue through at least the 65nm technology node and likely further [27].

Technology scaling has traditionally led to increased performance and reduced total energy consumption, but the tradeoff between leakage current growth and increased performance has become important and must be addressed. While low-power systems designers often use lower supply voltages to trade speed for power [17], for low-throughput applications such as sensor networks, designers should also consider leakage-performance tradeoffs between different process generations.

2.3 Low-Leakage Techniques

Several researchers have investigated architectural and circuit techniques to combat the significant increase of I_{leak} . While the majority of these techniques have been developed for performance-driven designs, they are also applicable to designs tuned for extremely low-power operation. Karnik et al. present a comprehensive overview of leakage reduction techniques and CAD challenges [15].

Dual- V_{th} process technologies allow designers to tune transistor performance and leakage at design time [31]. Threshold voltage can also be scaled by applying a back-body bias to the silicon substrate and wells. Both Intel and Hitachi have successfully demonstrated adaptive body biasing for low-power and high-performance processors [5, 8]. This work shows that by dynamically scaling threshold voltage, leakage current savings of up to $25\times$ can be achieved. This approach is effective for microprocessors that are designed to maintain maximum performance and adaptively change the threshold voltage to reduce leakage when the chip is inactive (sleep mode). While effective, indiscriminately increasing the bias voltage can lead to loss of state. Martin et al. conducted simulations that combined adaptive body biasing with dynamic voltage scaling (DVS) [20]. They found that the hybrid approach can reduce energy consumption by 23%-39% more than just DVS alone. Finally, both logic and memory circuits can achieve significant reduction in leakage current by gating the supply voltage at the cost of loss of state [24]

Leakage current is directly related to the current path from the supply to ground. Several researchers have shown that strategic transistor stacking can reduce leakage [21, 14]. Using input vector control to specify internal logic states has a similar effect, maximizing the number of series transistors that are off in the leakage path per gate [1].

Given our focus on low-throughput, energy-constrained applications, we focus on process and circuit-level leakage power reduction schemes, namely, the use of back-body biasing, Vdd-gating, and non-minimum length transistors.

3. A LEAKAGE AWARE ARCHITECTURE

3.1 Architectural Motivation and Goals

The example system employs an event-driven architecture designed for the regular nature of sensor network applications[13]. It includes a general-purpose microcontroller that spends most of the time in a low power state only awaking to handle irregular events such as system reprogramming. The event processor, which is a small state machine, handles all system interrupts and transfers data

between modularized slave components. The design goals of the example system are summarized below.

1. *Event-Driven Computation:* We seek to eliminate unnecessary event-processing software overhead by building a true event-driven hardware platform.
2. *Hardware Acceleration to Improve Performance and Power:* Our aim is a system composed of several components that are optimized for specific tasks. The intuition that drives this goal is that it is better to split the functionality of the system into several small components, each of which can be micro-managed for lower power consumption, as opposed to a monolithic computing engine that does not provide knobs for fine-grained power management.
3. *Exploiting Regularity of Operations within an Application:* We expect that specific hardware components will be able to handle regular events in an application input stream, thus avoiding the use of the general-purpose components, and minimizing energy consumption. Since irregular events occur infrequently, the penalty for using the general-purpose components of the system is justifiable.
4. *Optimization for a Particular Class of Applications:* Our architectural innovations aim to optimize the common-case behavior of monitoring applications for low-power, while still providing general-purpose processing capability for a broader class of applications.
5. *Modularity:* The system must be modular to allow different sets of hardware components to be combined into a larger system that is best suited to a particular type of application. A modular system architecture is easily extensible and is also well-suited to enabling and disabling blocks.
6. *Fine-grained Power Management Based on Computational Requirements:* One of the main themes driving our design is the possibility of configuring resource usage (for lower power consumption) of the sensor network devices on-the-fly according to computational demands. Fine-grained power management support at the architecture allows the designer to use advanced circuit techniques such as Vdd-gating and adaptive body bias to decrease leakage current.

3.2 Architecture Description

The system architecture is illustrated in Figure 1. There are two distinct divisions within the system in terms of the ability of the component to control the system bus. We refer to the components that have full control of the system address lines as *master* components and the remaining blocks that do not facilitate transfers on the databus as *slave* components. The system bus has three components – an interrupt bus, a data bus, and power control lines. The slaves respond to read or write requests from the master side of the databus, thus allowing the masters to read information content and control execution of the slaves. The two master devices consist of a general-purpose microcontroller and a small state machine, the event processor.

A key benefit of the modular design of our architecture is the ability to employ fine-grained power management of individual components (both masters and slaves). Selectively turning-off components, using Vdd-gating, allows to minimize leakage power. For example, the general-purpose microcontroller core could be relatively complex and power-hungry when active, but can be Vdd-gated most of the time when idling. The event processor handles

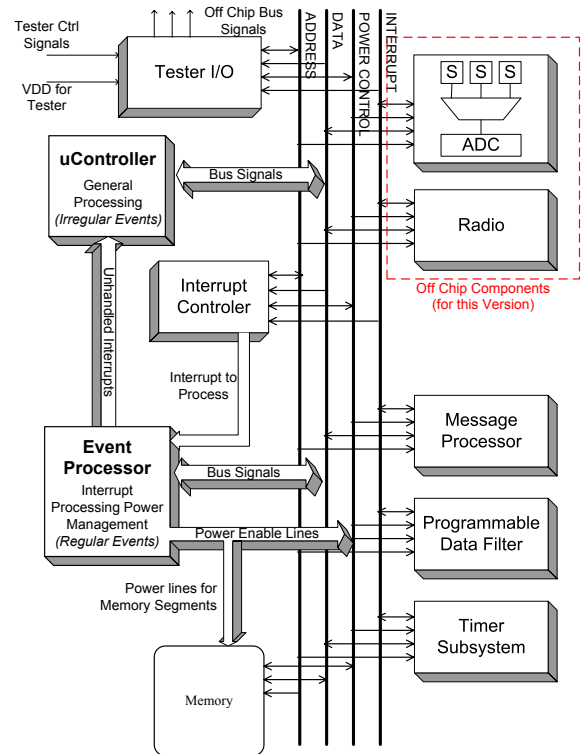


Figure 1: Block Diagram of System Architecture

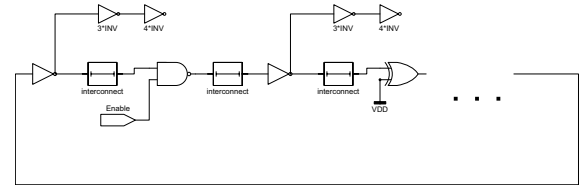


Figure 2: Test Circuit Used for Simulations. The circuit consists of an 11 stage ring oscillator made up of an assortment of logic gates. Interconnect was modeled between devices.

all interrupts, distributes tasks to slave devices, and wakes up the microcontroller only when necessary (rarely). A detailed architecture description is available in [13].

4. SIMULATION STUDY FOR TEST CIRCUITS

The goals of our simulation study are two fold – (1) to study active energy and leakage power tradeoffs across technology generations and (2) to investigate the impact of leakage current mitigation techniques in advanced process technology generations. In this analysis, we focus on threshold voltage scaling and channel length scaling.

4.1 Experimental Setup

Several HSPICE simulations were run to study the tradeoff between active energy and leakage power. The test circuit, illustrated in Figure 2, consists of an eleven-stage ring oscillator comprising a collection of static CMOS logic gates (inverter, NAND, NOR, XOR, etc.). This assortment of gates allowed us to take into account a reasonable mixture of different transistor configurations

and stacking effects when measuring the total power. One of the NAND gate inputs was used to disable the oscillator in order to measure leakage current. When simulating leakage current, sufficient time was allotted for the circuit to settle before collecting power consumption data.

We have found that models from different foundries follow slightly offset trends per technology generation. In order to ensure some level of consistency between technology generations, we use the Berkeley Predictive Technology Models (BPTM) [6, 4]. These predictive models have been found to be 10% accurate and are available from the 180nm technology node to the 45nm technology node. Because we are interested in the relative power consumption trends, BPTM models are sufficient for our study. We use BPTM models 180nm, 130nm, 100nm, and 70nm which all use the BSIM3 HSPICE model (BSIM4 models are not available for all technology nodes, so for consistency we rely on BSIM3). We also use the BPTM interconnect model in our test circuit to model interconnect loading effects across technology generations.

Our simulation environment allows us to sweep several different parameters of interest: technology generation, supply voltage, transistor size, and back-body voltage. Because we are mainly concerned with low throughput embedded applications we report data for room temperature simulations only and focus on trends across technology generations.

Figure 3 presents general trends observed across technology generations with nominal process corners and temperature. Figure 3-(a) clearly shows that leakage current increases significantly with newer process technologies and is sometimes even greater than the $5\times$ predicted by scaling theory.

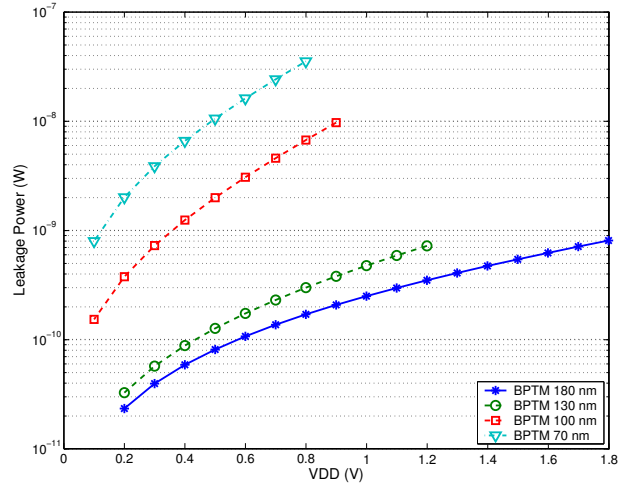
On the other hand, transistor scaling improves both active power and performance, evidenced by the reduction of energy-delay product. These opposing trends in leakage power and power efficiency (EDP) motivate further detailed analysis to understand how the choice of different technology nodes affect low-throughput applications. Looking closely at Figure 3-(b) one notices that for low voltages (roughly less than double the threshold voltage), EDP increases drastically. This is due to the low current driving capability of transistors operating in the subthreshold region. While subthreshold designs can lead to low-power operation [30], circuit delays can vary significantly given the exponential dependence on voltage. Hence, significant expansion of timing margins or asynchronous operation are required for successful operation. To somewhat simplify design and analysis, we conservatively target the supply voltage with the minimum EDP. For low-throughput applications, this leads to a clock-gated design, which computes synchronously during active periods and then consumes leakage current when idle, instead of running at the lowest operating voltage and frequency for a given circuit frequency.

Figure 3-(c) plots the operating frequency of the test circuit observed across technology generations and supply voltages. As can be expected, these results show an increase in operating frequency for more advanced process technology nodes. For a given supply voltage and technology we allowed 1ms for the circuit to complete 4 successful oscillations or else the simulation was stopped and the data was not included in the plot.

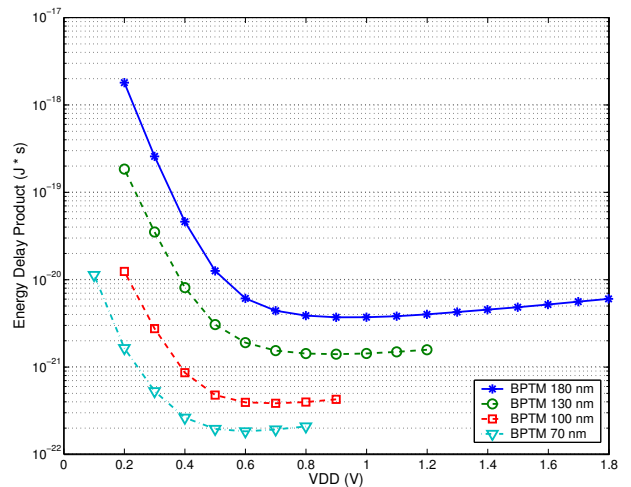
4.2 V_{th} Scaling

As discussed in Section 2, scaling the threshold voltage by applying a reverse body bias or using high- V_{th} transistors is a standard technique that is used when the primary design goal is high performance with low standby power consumption.

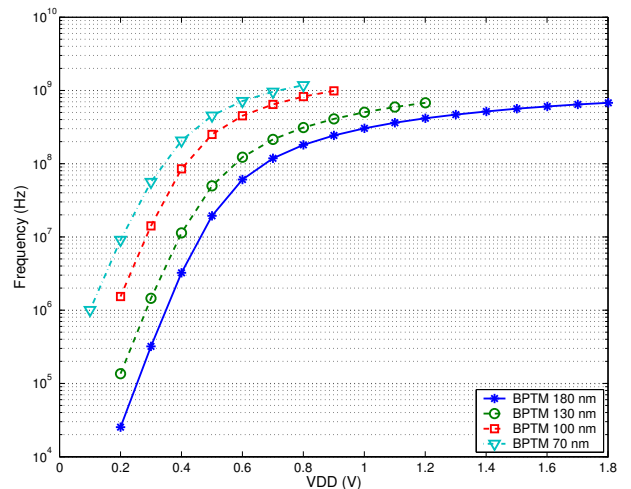
Applying a bias voltage on the silicon substrate to dynamically scale threshold voltage has been used effectively to decrease leak-



(a) Leakage power

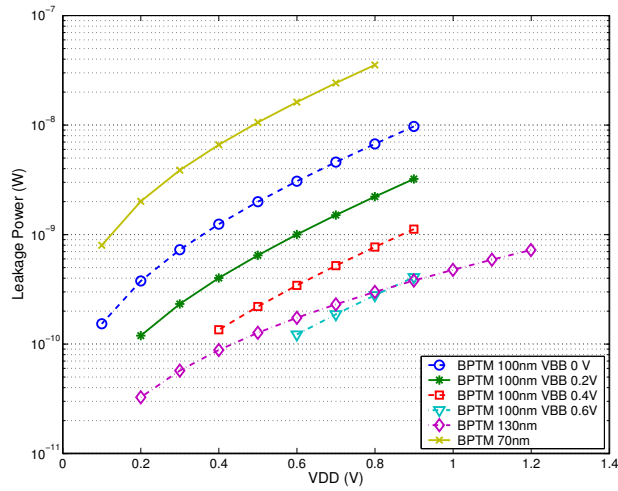


(b) Active Energy

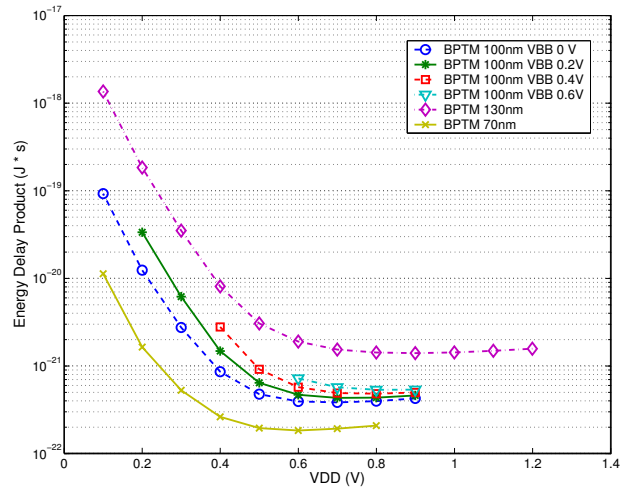


(c) Frequency

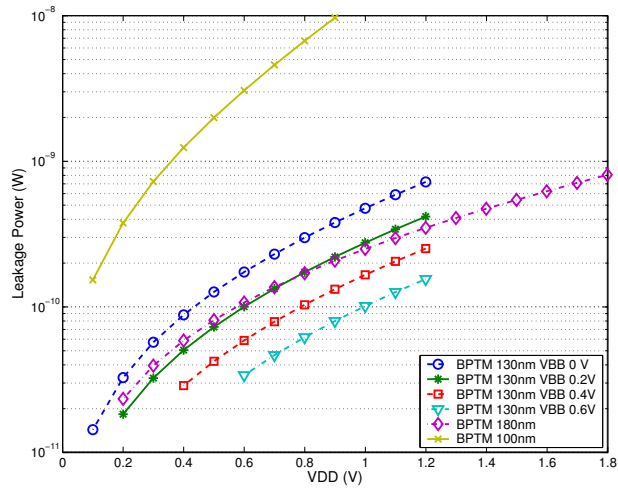
Figure 3: Leakage Power, EDP, and Frequency Across all Technologies Each line indicates a technology node from 180nm to 70nm. Supply voltage is on the X-axis which was swept from 0.1V to the max VDD specific to the process. Temperature is 20°C and all transistors are minimum size.



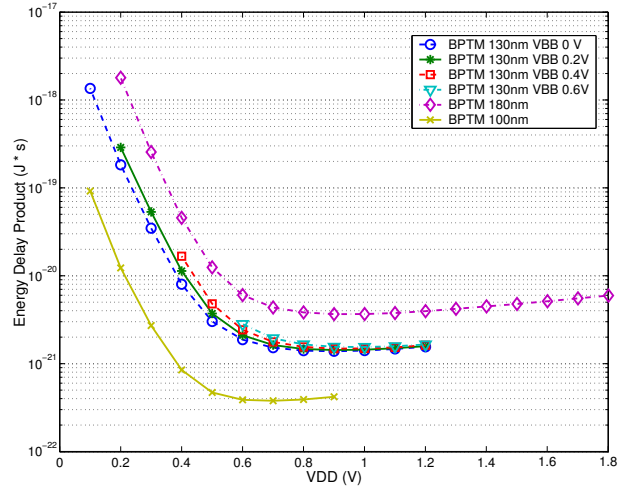
(a) Back Body Bias Leakage Power - 100nm



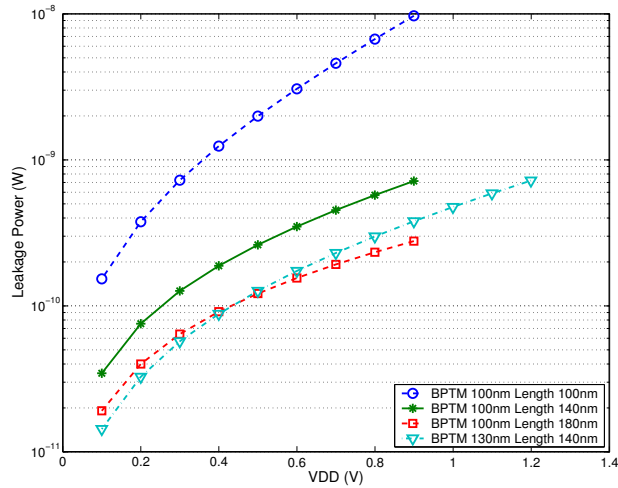
(b) Back Body Bias EDP - 100nm



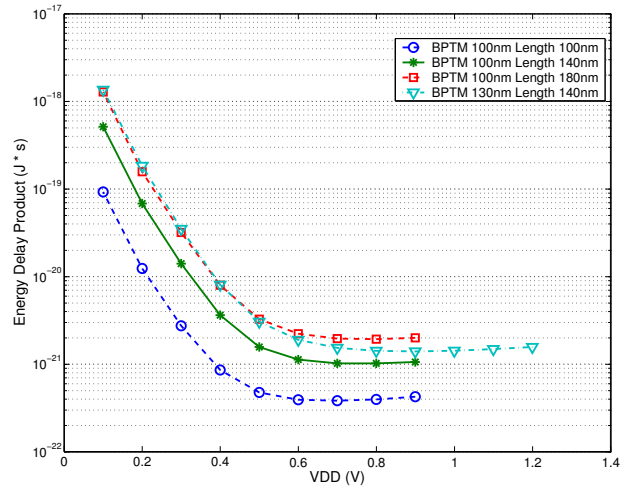
(c) Back Body Bias Leakage Power - 130 nm



(d) Back Body Bias EDP - 130 nm



(e) Longer Channel Length Leakage Power - 100 nm



(f) Longer Channel Length EDP - 100 nm

Figure 4: Power Consumption With Back Body Bias and Longer Channel Lengths. Plots are displayed with back-body biasing applied to both 100nm and 130nm technology nodes. Temperature is 20°C and all transistors are minimum size.

age current. To illustrate the effectiveness of this technique, simulation results with back-body biasing applied to both 100nm and 130nm technology nodes are presented. As shown in Figures 4-(a) and 4-(c), back-body biasing decreases the overall leakage current of our test circuit. It is interesting to note that applying a back-body bias on the 100nm technology results in a decrease of leakage current, but not less than the total leakage current of 130nm circuitry without biasing. On the other hand, applying a back-body bias to 130nm circuitry reduces leakage current below that of 180nm circuitry without biasing, attributable to the already small difference in leakage current between the two technology nodes.

In addition to leakage power, we consider the effects of V_{th} scaling on active energy and performance. Figures 4-(b) and 4-(d) show that the resulting increase in threshold voltage leads to decreases in performance and higher energy-delay product. Because of the smaller parasitic capacitances (device and interconnect), the delay of circuitry with back-body biasing is still faster than circuitry in an older generation without biasing, resulting in a lower energy-delay product.

From these simulation results we conclude that back-body biasing is much more effective for circuitry in the 130nm node than the 100nm node, where leakage current reduction is much more pronounced. If leakage power is the primary concern, as in low-throughput applications, an older technology generation can be used and back-body biasing can be applied to further reduce leakage. It is also important to mention that back-body biasing requires special-purpose circuitry and comes with a power penalty, which has been ignored.

4.3 Channel Length Scaling

By increasing the drawn channel length of transistors, it is possible to reduce leakage current. This technique, like selection of high- V_{th} transistors, must be performed statically at design time, and the results are representative of similar techniques such as transistor stacking and selective input vector activation.

Figures 4-(e) and 4-(f) present results for circuitry simulated with several different channel lengths. Increasing channel length larger than the native length of the older technology generation drastically decreases leakage power. We also observe that when channel length increases to the size of the older generation its active power characteristics follow that of the older technology. The slight difference in energy and leakage power can be explained by the differences in the minimum widths of the transistors as well as gate oxide thickness and other process specific parameters.

While scaling transistor length does decrease leakage power, the loss of active energy means that little benefit is gained from going to a more advanced process technology. Increasing transistor length could be a useful technique when older processes are not available or a designer wishes to construct different circuits for high performance and low power on the same die.

5. MODELING ARCHITECTURE ACROSS PROCESS TECHNOLOGIES

The analysis of a simple ring oscillator reveals the influence of process technology selection and leakage control techniques on both active power and leakage power consumption. However, the total energy consumed by a system is dependent on workload, power supply voltage, system architecture, and the application of energy reduction techniques. In this section we describe a power model of our system that allows us to study the effect of different architecture and circuit techniques across process technology nodes. The next section describes the results of this study.

5.1 Architecture Model

We developed a power model for our system by using the results from Section 4 combined with power estimates for synthesized RTL blocks in our system architecture. We consider sensor network applications that are periodic in nature and can be modeled as a combination of discrete events which we refer to as a *task*. For example, a typical task would consist of the following series of events: an internal timer fires, the sensor node collects a sensor sample, the filtering block performs local data processing, the message processor prepares a radio message with relevant data, and the packet is aggregated and possibly transmitted as a radio message. After completing the task, the node idles until the next timer interrupt indicates that a new task must be executed.

We model the total energy consumption during one second of operation as the energy consumed in a task plus the energy consumed between tasks:

$$E_{total} = N * E_{task} + (F/C_{task} - N) * E_{inter-task}$$

Where F is the main clock frequency of the system (which is a free variable in our analysis), C_{task} is the number of clock cycles per task for a given application (for our test workload, this is 131), and N is the number of tasks executed in one second. For typical environmental monitoring applications N is fairly small and depends on the characteristics of the phenomena being measured. A survey of the literature indicates that N is often in the range of 0.02 to 100 in actual deployments [19, 32].

The energy consumption during a task and between tasks is equal to the sum of the energy consumption of each hardware block. Each block can either be active, idle (leaking), or leakage-managed (via Vdd-gating or RBB) depending on the behavior of the application.

$$E_{task} = F/C_{task} * \sum_{i=1}^{all.blocks} \alpha_i * P_{a_i} + g * P_{gate_i} + l * P_{leak_i}$$

$$E_{inter-task} = F/C_{task} * \sum_{i=1}^{all.blocks} A_i * P_{a_i} + G * P_{gate_i} + L * P_{leak_i}$$

Where α, g, l are activity factors within a task for each block. We gather these factors from application traces with our Verilog model. To denote the activity factors between tasks we use A, G, L . We describe the power consumption of each block with the variables P_a, P_{gate}, P_{leak} which are functions of process technology, supply voltage, and frequency. E_{task} is the sum of the energy consumed active, idle and gated modes for each block while the system is computing a task. The mode of operation of each block depends on the application behavior. $E_{inter-task}$ sums up the total energy while the system is between tasks.

We used a standard cell library and Synopsys Design compiler to synthesize each block and generate power estimates for delay, active power, and leakage power in a 130nm process technology. We then used the scaling factors from our ring oscillator simulation study to scale these baseline numbers across process technologies. Table 1 presents the scaling factors obtained from these simulations and how these factors compare to established technology scaling trends. We observed that active power scales 0.49 per technology node, while leakage power increases 6x per technology generation. Delay scales 0.84.

	Constant Field	Constant Field
Scaling Factor	Theory [2]	Simulation
Active Power	0.49	0.49
Leakage Power	5	6.0
Delay	0.7	0.84

Table 1: Scaling Factors *From theory and simulation data*

	α	g	l	A	G	L
	In Task			Inter-Task		
Component	active	gated	idle	Active	Gated	Idle
uProc	0.00	1.00	0.00	0.00	1.00	0.00
Timer	0.25	0.75	0.00	0.25	0.75	0.00
Filter	0.58	0.42	0.00	0.00	1.00	0.00
Message Proc	0.80	0.20	0.00	0.00	1.00	0.00
SRAM	0.21	0.75	0.04	0.00	0.75	0.25
Interrupt Ctrl	0.05	0.00	0.95	0.00	0.00	1.00
Event Processor	0.92	0.00	0.07	0.10	0.00	0.90

Table 2: Activity Ratios for Our Test Application

In order to scale frequency and supply voltage, we use the data for each process technology presented in Section 4. We first interpolate this data for 1mV steps and then use the normalized power, voltage, and frequency scaling relationships to scale our baseline power and frequency data.

Finally, we perform detailed simulations to estimate the energy savings available from Vdd-gating. We sweep the size of the gating transistor for one of our synthesized blocks and find that Vdd-gating can reduce overall leakage current by a range of 10-400 \times . We find that by carefully selecting the gating transistor size, savings of 100 \times can be achieved without an appreciable impact on delay which agrees with similar results reported in the literature [24]. Therefore we scale the leakage current of a block by 1/100 if it is eligible to be Vdd-gated. This model does not take into account the cost of powering up and down a block, or the active power overhead of the Vdd-gating transistor.

5.2 Test Application

Our test application is typical of a wide range of environmental monitoring sensor network workloads. A task in our test workload consists of the following events: a timer fires, a data sample is collected, a filtering operation is performed, and a message is prepared and transmitted. This application has been ported to our architecture and was used in the verification stages of our system.

We collect block-level activity traces from the Verilog model of our system. These activity factors are enumerated in Table 2. Notice that many slave blocks can be Vdd-gated both within a task and between tasks. Because the Event Processor and Interrupt Controller blocks need to be able to process incoming interrupts they are not Vdd-gated in our base architecture. The Timer block includes 4 distinct timers and only one timer is required for this application which explains why 75% of the block is gated. Likewise the SRAM architecture includes 4 separate banks that can be individually gated (although typically some state will need to be preserved between tasks).

Using our power model and the application activity traces we are able to study the energy consumption of our system using various low energy techniques across process technologies. The following section describes these results.

6. RESULTS OF SYSTEM ANALYSIS

This section presents the evaluation of different architecture and circuit techniques across process technology nodes using the power model and test application presented in Section 5. First we present results for our baseline architecture and discuss the limits of voltage scaling. Then we model different low power techniques such as Vdd-gating, using multiple clocks, and reverse body-bias for memory circuits and show how these techniques impact the total energy consumption of the system. Finally we summarize the contribution of each of these techniques and discuss our observations.

All of our plots sweep frequency from the minimum frequency defined by the cycles per task multiplied by the total tasks for one second ($C_{task} * N$) up to 100 MHz (due to characteristics of our applications, we do not consider frequencies beyond 100 MHz). At each frequency step, the minimum voltage that met the frequency target was selected. Using this voltage and frequency pair, the corresponding values for P_a , P_{gate} , P_{leak} were selected from the dataset described in Section 5 for each block. Total energy for 1 second of operation is then calculated. In the majority of the plots we fix the number of tasks (N) at 100 to reflect the upper bound of the application requirements.

6.1 Baseline Architecture

Figure 5 presents results for our baseline architecture without any energy-reduction techniques except standard clock gating. In both plots we show how process technology impacts both energy consumption and the selection of supply voltage. Figure 5-(a) shows that 180nm provides the lowest energy consumption for low system frequencies but as frequency is increased the more advanced process technologies have lower energy consumption. This is due to an increase in inter-task active power (e.g. blocks that continue to run during the idle period). Specifically, Table 2 reveals that the timer component is required to be active between tasks so that the system knows when to begin the execution of the next task. Therefore as the overall clock frequency of the system increases, the active power between tasks will dominate the overall energy consumption of the system.

In this analysis, we utilize the lowest supply voltage that will meet the performance requirements of the system. Figure 5-(b) shows the voltage frequency relationship for the baseline architecture across process technologies. The voltage and frequency relationships were scaled from ring-oscillator simulations with a minimum oscillation frequency of 10KHz. We see from this figure that the minimum supply voltage is either 0.2 V or 0.1 V. While, static CMOS logic circuits can operate at these low voltages it has been shown that traditional SRAM designs are limited by their static noise margins to a supply voltage of around two times the threshold voltage depending on topology [3]. While it is possible to use latch-mux based memory designs in subthreshold, these designs have the drawback of much larger cells [30]. Calhoun et al. [3] show that it is possible to build a low voltage 10T SRAM that operates down to 350 mV in the 65nm technology node. However, robustness to process variability may limit the practical voltage limit with this approach to some degree. Given these robustness and memory density issues, designers may opt for more traditional 6T SRAM designs used in many standard cell libraries and memory generators; in this case, the supply voltage will be limited to a minimum of $2*V_t$. For all subsequent plots we limit the minimum supply voltage to $V_{tP} + V_{tN}$ with traditional 6T SRAM designs.

Figure 6-(a)+(b) presents the baseline architecture with the power supply voltage limited as described. Notice that the trends are consistent with the previous plot, but the overall energy consumption has increased.

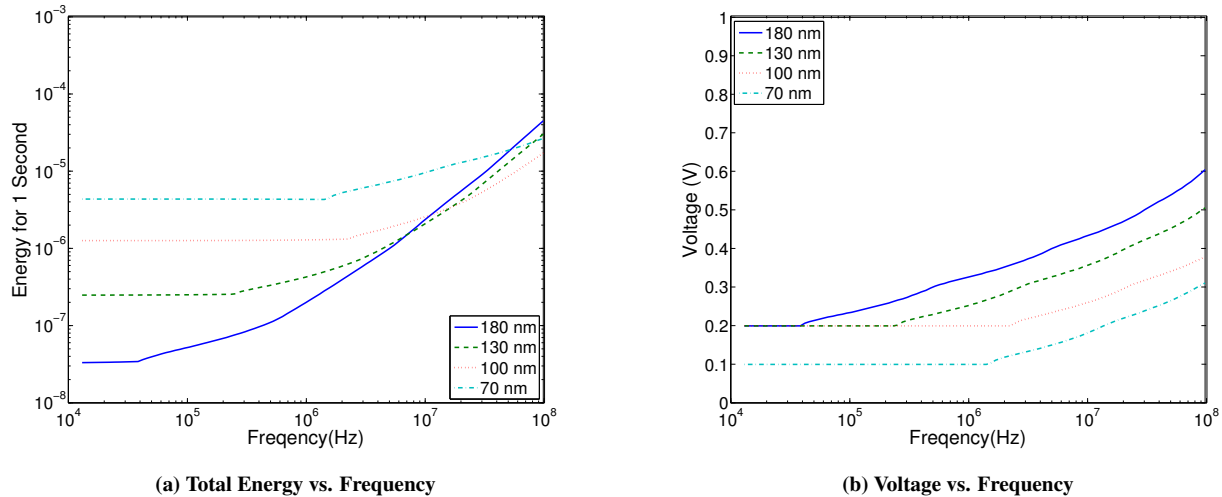


Figure 5: Results for Baseline Architecture. Performance target of $N=100$ sense and transmit tasks.

6.2 Effect of Energy Reduction Techniques

Our modular architecture was designed to support a variety of energy reduction techniques. First we model Vdd-gating on the block level as implemented by our system. Figure 6-(c) presents results with Vdd-gating included across clock frequencies and process technologies. Notice that for leakage-dominant technologies total energy has been significantly reduced compared with the baseline architecture in Figure 6-(a). However, at low frequencies the system will have little idle time available for Vdd-gating the block. This explains why for more advanced process technologies at lower frequencies the curve slopes upward. At these low frequencies, the workload's total energy is dominated by leakage power so increasing the amount of time the system can be Vdd-gated will decrease energy consumption. However, just like in the baseline architecture if the system clock frequency is increased significantly the inter-task active components such as the timer and event processor will dominate the total energy consumption.

To decrease the inter-task energy consumption we model a system with two clocks, one higher frequency system clock and one low frequency clock with frequency equal to N . This is similar to the techniques employed by low power processors such as the TI MSP-430[29]. We also note that the entire event processor and interrupt controller do not need to be powered on to receive an interrupt. In reality, a small state machine could remain powered on while the event processor and interrupt controller are Vdd-gated. We model this architecture feature by gating the leakage power consumed by these components. The result of these techniques combined with Vdd-gating is shown in Figure 6-(d). Notice how the use of two-clocks has made the inter-task active energy frequency invariant. The use of wakeup circuits only has a noticeable impact on the more advanced process technologies dominated by leakage. For this system configuration, the highest clock-frequency for the supply voltage of $Vt_P + Vt_N$ is the energy-optimum operating point. This result is somewhat non-intuitive, but in this architecture running at the highest frequency maximizes the amount of time that blocks can be Vdd-gated.

The SRAM banks that cannot be gated still consume significant leakage energy. Researchers have shown that it is possible to use reverse-body-bias (RBB) to dynamically increase threshold voltage thereby decreasing subthreshold leakage energy of the block. As

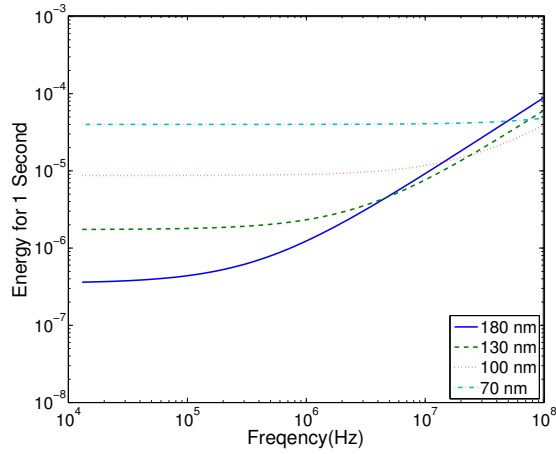
described in the literature it is possible to decrease the leakage current using RBB by a factor of $5.9 \times [10]$. It would also be possible to achieve a similar effect by reducing the power supply of the SRAM; however, we are already using a significantly reduced power supply so this provides little additional benefit. We model the reduction of leakage current using RBB by decreasing the inter-task leakage current of the ungated blocks by a factor of $5.9 \times$. Figure 6-(e) shows that the use of RBB further reduces the total energy consumption of the system for the more advanced technologies dominated by leakage power. We see that the 130nm technology node has nearly the same energy consumption as the 180nm node once all of the energy reduction techniques have been included. However, it is obvious that the 70nm and 100nm technology nodes are still not competitive from an energy-efficiency standpoint despite the large reductions in energy provided by these techniques.

Figure 6-(f) details how the results vary as the number of tasks per second (N) is swept. The plot is consistent with the previous plots for low values of N , but as N increases the total energy consumption of the system is dominated by active power and the more advanced process technologies are superior. However, as most sensor network deployments have values of N that are less than 100, advanced technologies are only likely to be useful if sensor nodes begin to perform additional local computation or more complex tasks.

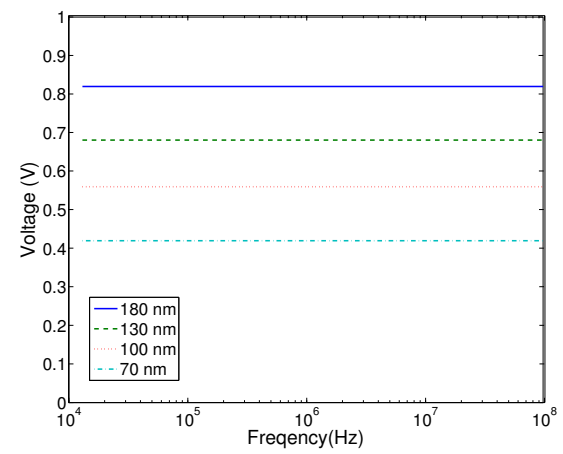
6.3 Summary and Discussion

Figure 7 provides a summary of the energy reduction techniques that we have modeled. Each bar represents the lowest total energy observed for a particular system configuration and process technology. Figure 7 (a) presents the total energy consumption. Generally the more advanced process technologies have higher energy consumption for low-throughput applications because total energy consumption is dominated by leakage current. The circuit and architecture techniques that we model can reduce the total energy consumption of the system by more than an order of magnitude. As expected, the leakage-dominant technologies show the greatest energy reduction.

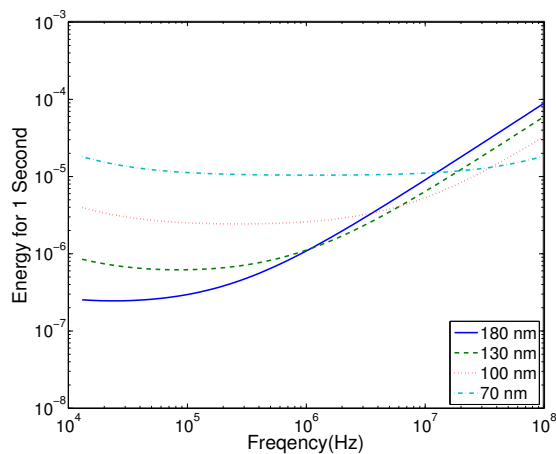
Figure 7 (b) provides a percent breakdown of the sources of energy consumption in the system. For the baseline configuration, operating with the lowest frequency and voltage minimizes energy



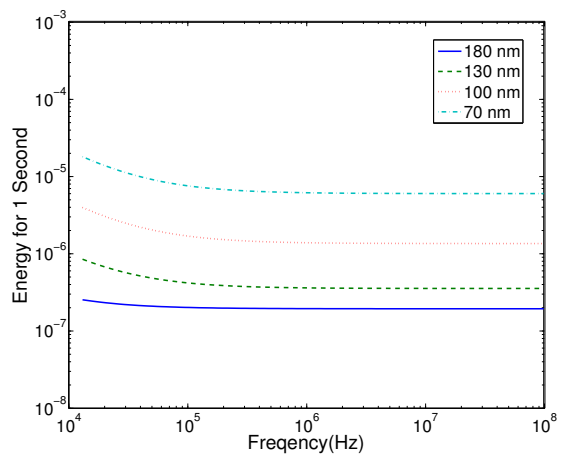
(a) Baseline Architecture



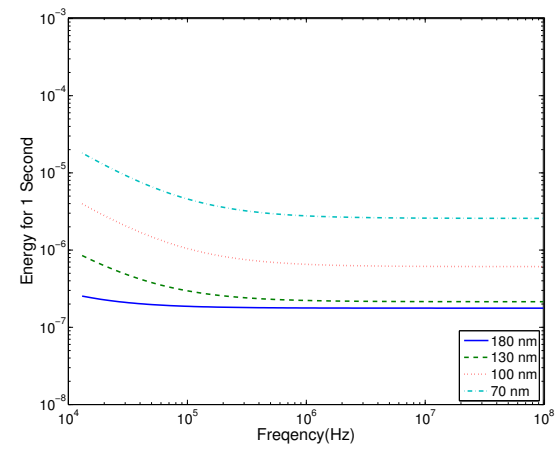
(b) Baseline Architecture - Voltage vs. Frequency



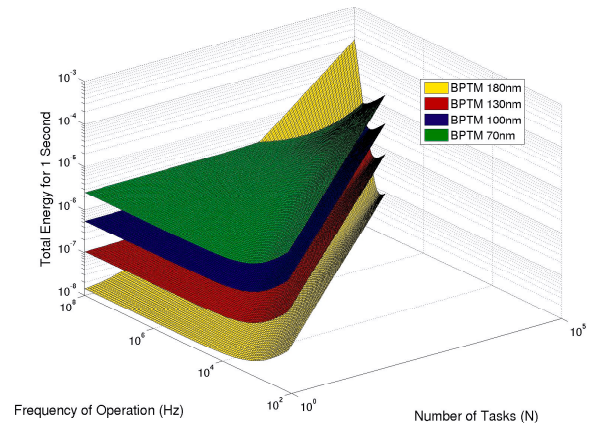
(c) VDD gating



(d) VDD gating, two clocks, and wakeup circuits



(e) Including RBB for SRAM and techniques from (b)



(f) 3D Summary of (c) sweeping N

Figure 6: Effect of Energy Reduction Techniques on Total Energy Consumption of the Architecture Across Process Technologies. Power Supply voltage is limited to $Vt_P + Vt_N$ and the number of tasks per second is 100.

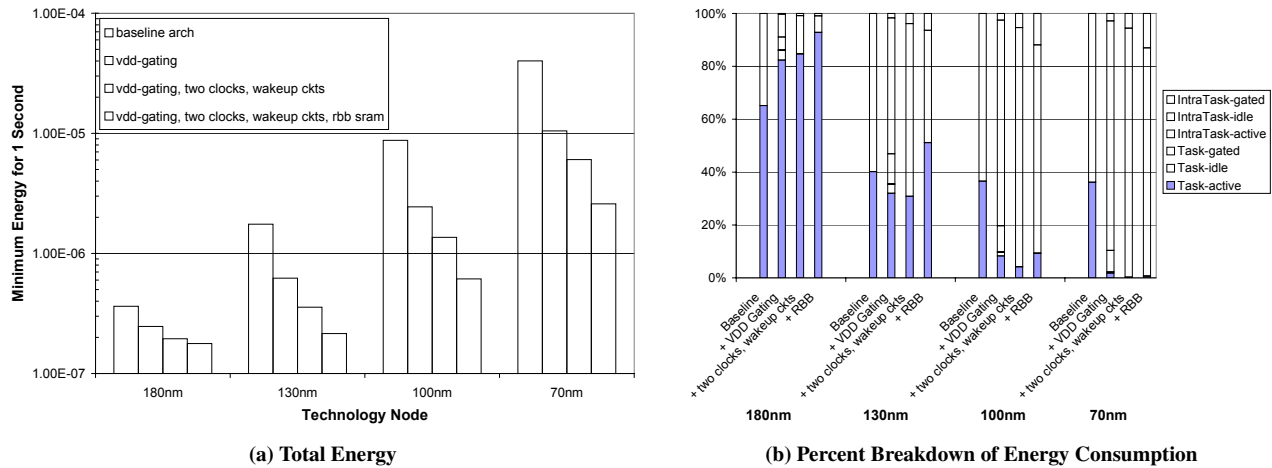


Figure 7: Summary of Energy Reduction Techniques Across Process Technologies Each bar represents the minimum energy calculated for a particular architecture configuration and process technology. Both the total energy consumption and a percentage breakdown of the source of energy consumption are included.

consumption. Therefore most of the energy is consumed during task execution. Once VDD-gating is included it is more energy efficient to run at higher clock frequencies and sleep during inter-task periods. For these configurations energy is consumed primarily through leakage. For the 180nm process technology node, active energy dominates while the energy consumption in the 70nm node is nearly all leakage energy (idle + gated). Across all process technologies idle and gated energy within a task is small due to the low throughput requirements of sensor network workloads. Therefore when using more advanced process technologies, architects and circuit designers should focus design efforts on reducing idle energy. The contribution of intra-task active energy can be significantly reduced by including two clocks and wakeup circuits. Adding reverse body bias to the SRAM further reduces the contribution of intra-task idle energy.

From the analysis above we can enumerate several key *lessons learned* that can guide the development of future sensor node platforms.

- **Reduce Supply Voltage.** It is important to set Vdd as low as possible for reliable operation at the target frequency; standard SRAMs have scaling limits that should be followed.
- **Maximize Vdd-gating time.** Extending the amount of time to Vdd-gate a block can decrease leakage and overall energy; raising clock frequency to maximize this period is a non-intuitive choice that can yield benefits.
- **Provide architectural support for Vdd-gating.** More fine-grain gating provides more opportunities for leakage energy reduction.
- **Minimize inter-task energy consumption.** If a block needs to compute between tasks (e.g. interrupt controllers and timers), multiple clocks can minimize active energy.
- **Process technology choice is an energy reduction technique.** For low throughput applications where leakage energy dominates, the choice of process technology can have a significant impact on total energy consumption.

- **Reduce the energy consumption of blocks that cannot be gated.** Using RBB for state-preserving SRAMs further reduces the energy consumption of the system.

There are other techniques such as using high-Vt transistors and longer channel lengths that we have investigated. These techniques do result in a significant reduction in leakage current, but they are static in nature and will not respond to changing workload and application requirements. Our future work will incorporate additional energy reduction techniques across process technologies and enhance our results with physical measurements of our system.

7. CONCLUSIONS

The emerging space of low-throughput sensor network applications requires significant architecture and circuit support to cope with the increased leakage currents of advanced nano-scale process technologies.

This paper performs extensive HSPICE simulations to quantify leakage power, active energy consumption, and delay across different process technology nodes. This analysis was used to guide the study of our system architecture. We have evaluated several energy reduction techniques such as Vdd-gating and adaptive body bias for a low-power system architecture across a variety of process technologies.

The results are clear; designers of future systems for low-throughput applications need to balance active and leakage energy consumption across all levels of the design space from the architecture down to the circuits and the selection of a process technology node.

Acknowledgments

This work is supported by NSF grants CNS-0330244, CCF-0048313 (CAREER), Intel, and IBM. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, Intel or IBM.

References

- [1] Afshin Abdollahi, Farzan Fallah, and Massoud Pedram. Leakage current reduction in CMOS VLSI circuits by input vector control. *IEEE Trans. VLSI*, 12(2):140–154, Feb 2004.

- [2] Shekhar Borkar. Design challenges of technology scaling. *IEEE Micro*, pages 23–29, July-Aug 1999.
- [3] Benton H. Calhoun and Anantha Chandrakasan. A 256kb sub-threshold SRAM in 65nm CMOS. In *IEEE International Solid-State Circuits Conference (ISSCC)*, February 2006.
- [4] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu. New paradigm of predictive MOSFET and interconnect modeling for early circuit design. In *Proc. IEEE CICC*, June 2000.
- [5] Lawrence T. Clark, Shay Demmons, Neil Deutscher, and Franco Ricci. Standby power management for a 0.18 μ m microprocessor. In *Proc. ISLPED*, Aug 2002.
- [6] Device Research Group EECS, UC Berkeley. BSIM homepage official site. <http://www-device.eecs.berkeley.edu/~bsim/>.
- [7] Virantha Ekanayake, IV Clinton Kelly, and Rajit Manohar. An ultra low-power processor for sensor networks. In *Proc. ASPLOS*, Oct 2004.
- [8] Hiroyuki Mizuno et al. An 18- μ A standby current 1.8-V 200-MHz microprocessor with self-substrate-biased data-retention mode. *IEEE Journal of Solid-State Circuits*, 34(11):1492–1500, 1999.
- [9] Nam Sung Kim et al. Leakage current: Moore’s law meets static power. *IEEE Computer*, pages 68–75, Dec 2003.
- [10] Krisztian Flautner, Nam Sung Kim, Steve Martin, David Blaauw, and Trevor Mudge. Drowsy caches: Simple techniques for reducing leakage power. In *The 29th Annual International Symposium on Computer Architecture (ISCA)*, May 2002.
- [11] David J. Frank, Robert H. Dennard, Edward Dowak, Paul M. Solomon, Yuan Taur, and Hon-Sum Philip Wong. Device scaling limits of Si MOSFETs and their application dependencies. *Proceedings of the IEEE*, 89(3):259–288, March 2001.
- [12] Thaddeus R. F. Fulford-Jones, Gu-Yeon Wei, and Matt Welsh. A Portable, Low-Power, Wireless Two-Lead EKG System. In *Proceedings of the 26th IEEE EMBS Annual International Conference*, San Francisco, CA, Sept 2004.
- [13] Mark Hempstead, Nikhil Tripathi, Patrick Mauro, Gu-Yeon Wei, and David Brooks. An ultra low power system architecture for sensor network applications. In *The 32nd Annual International Symposium on Computer Architecture (ISCA)*, June 2005.
- [14] Mark C. Johnson, Dinesh Somasekhar, and Kaushik Roy. Leakage control with efficient use of transistor stacks in single threshold CMOS. In *Proc. 36th ACM/IEEE Conference on Design Automation*, June 1999.
- [15] Tanay Karnik, Shekhar Borkar, and Vivek De. Sub-90nm technologies—challenges and opportunities for cad. In *International Conference on Computer-Aided Design (ICCAD)*, November 2002.
- [16] Dongwoo Lee, David Blaauw, and Dennis Sylvester. Gate oxide leakage current analysis and reduction for VLSI circuits. *IEEE Trans. VLSI*, 12(2):155–166, Feb 2004.
- [17] Dake Liu and Christer Svensson. Trading speed for low power by choice of supply and threshold voltages. *IEEE Journal of Solid-State Circuits*, pages 10–17, January 1993.
- [18] Konrad Lorincz, David Malan, Thaddeus R. F. Fulford-Jones, Alan Nawoj, Antony Clavel, Victor Shnayder, Geoff Mainland, Steve Moulton, and Matt Welsh. Sensor networks for emergency response: Challenges and opportunities. *IEEE Pervasive Computing*, Oct-Dec 2004.
- [19] Alan Mainwaring, Joseph Polastre, Robert Szewczyk, David Culler, and John Anderson. Wireless sensor networks for habitat monitoring. In *ACM International Workshop on Wireless Sensor Networks and Applications (WSNA’02)*, Atlanta, GA, September 2002.
- [20] Steven M. Martin, Krisztian Flautner, Trevor Mudge, and David Blaauw. Combined dynamic voltage scaling and adaptive body biasing for lower power microprocessors under dynamic workloads. In *International Conference on Computer-Aided Design (ICCAD)*, November 2002.
- [21] Siva Narendra, Vivek De, Shekhar Borkar, Dimitri Antoniadis, and Anantha P. Chandrakasan. Full-chip sub-threshold leakage power prediction model for sub-0.18 μ m CMOS. In *ISLPED*, Aug 2002.
- [22] Leyla Nazhandali, Michael Minuth, Bo Zhai, Javin Olson, Scott Hanson, Todd Austin, and David Blaauw. A second-generation sensor network processor with application-driven memory optimizations and out-of-order execution. In *ACM/IEEE International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES)*, September 2005.
- [23] Leyla Nazhandali, Anna Reeves, Michael Minuth, Ryan Helfand, Javin Olson, Bo Zhai, Sanjay Pant, Todd Austin, and David Blaauw. Energy optimization of subthreshold-voltage sensor processors. In *The 32nd Annual International Symposium on Computer Architecture (ISCA)*, June 2005.
- [24] Michael Powell, Se-Hyun Yang, Babak Falsafi, Kaushik Roy, and T.N. Vijaykumar. Gated-vdd: A circuit technique to reduce leakage in deep-submicron cache memories. In *International Symposium on Low Power Electronics and Design (ISLPED)*, June 2000.
- [25] Rajeev Rao, Ashish Strivastava, David Blaauw, and Dennis Sylvester. Statistical analysis of subthreshold leakage current for VLSI circuits. *IEEE Trans. VLSI*, 12(2):131–139, Feb 2004.
- [26] Shad Roundy, Paul K. Wright, and Jan Rabaey. A study of low level vibrations as a power source for wireless sensor nodes. *Computer Communications*, 26(1):1131–1144, July 2003.
- [27] Semiconductor Industry Assoc. International technology roadmap for semiconductors, 2004 update. <http://www.itrs.net/Common/2004Update/2004Update.htm>.
- [28] R. Szewczyk, J. Polastre, A. Mainwaring, and D. Culler. Lessons from a sensor network expedition. In *Proc. the First European Workshop on Wireless Sensor Networks (EWSN)*, January 2004.
- [29] Texas Instruments. TI MSP430F149 Ultra-Low Power Microcontroller. <http://www.ti.com>.
- [30] Alice Wang and Anantha Chandrakasan. A 180mV FFT Processor Using Subthreshold Circuit Techniques. In *Proc. ISSCC*, Jan 2004.
- [31] L. Wei, Z. Chen, K. Roy, M. Johnson, Y. Ye, and V. De. Design and optimization of dual threshold circuits for low-voltage low-power applications. *IEEE Trans. VLSI*, 7(1):16–24, Mar 1999.
- [32] Geoffrey Werner-Allen, Jeff Johnson, Mario Ruiz, Jonathan Less, and Matt Welsh. Monitoring volcanic eruptions with a wireless sensor network. In *Proceedings of the Second European Workshop on Wireless Sensor Networks (EWSN)*, January 2005.